

The returns from arthritis research

A report prepared by RAND Europe for the Arthritis Research Campaign (arc)



20041110 046
HERG



RAND
EUROPE

Health Research & Evidence Group

 0870 850 5000
www.arc.org.uk
Committed to curing arthritis

The research described in this report was prepared for and funded by the Arthritis Research Campaign (**arc**).

Cover photos illustrate some examples of Research payback:

(left) **arc**-funded research demonstrates that the use of aspirin and heparin for pregnant women with APS increases the live birth rate by 40% compared to the use of aspirin alone and by 60% compared to no treatment at all.

Photo © Paul Preacher

(right) **arc**-funded research demonstrates that OA of the hip is a condition that arises through the interaction of a predisposition to the disease, and mechanical insults to the hip. It is now accepted by the Industrial Injuries Advisory Council that hip OA in farmers is a prescribed disease, meaning that farmers have become eligible for appropriate compensation payments.

Photo © Kate Bellis

(main picture) Basic research at **arc**'s flagship Institute led to the development of a new category of drugs. Hundreds of thousands of patients worldwide have been treated, of whom 70% experience a significant improvement in their health.

ISBN 0-8330-3687-4

RAND Europe is a nonprofit research organisation providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

© Copyright 2004 Arthritis Research Campaign

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from Arthritis Research Campaign.

This report can be downloaded for personal use at no charge from
<http://www.rand.org/publications/MG/MG251/>

ISBN: 0833036874

Technical rpt #: RAND/MG-251-ARC

Order No./Price: \$20.00 paperback

Title: The returns from arthritis research volume 1 : approach, analysis and recommendations / Steven Wooding ... [et al].

Electronic access: Online access

<http://www.rand.org/publications/MG/MG251/>

Publication info: Santa Monica, CA : RAND, 2004.

Physical description: xxi, 105 p. : ill. (some col.) ; 30 cm.

Note: "A report prepared by RAND Europe for the Arthritis Research Campaign (arc)" --Cover.

Note: "RAND Europe."

Note: Includes bibliographical references.

Abstract: There is increasing pressure for research funders to demonstrate, and seek to maximise, the payback from the research they fund. This report, prepared for and funded by the Arthritis Research Campaign (arc), presents the results of an evaluation of 16 research grants awarded by arc in the early 1990s. The main objective was to develop a system for evaluating arthritis research, with a view to allowing arc to stimulate and manage the exploitation of research advances so that they translate into outcomes of practical benefit to people with arthritis. The report presents a framework that conceptualises the relationship between research inputs, process, output and outcomes. Using this framework, we catalogue a diverse range of research output and outcomes arising from these 16 grants and make a series of quantitative and qualitative assessments comparing, for example, payback from project grants versus programme grants. In conclusion, we make six observations: --There is a diversity of research payback. --The researcher is the key driver of research translation. --Short, focused project grants seem to provide value for money. --Intended and unintended flexibility in funding is used advantageously. --Referees' contributions to the peer-review process are of variable benefit. --The payback framework could be operationalised and embedded by arc. The companion Volume 2 is a collection of the case studies. These case studies all follow a similar format based on the conceptual model and provide a rich and detailed narrative on the payback of each research grant.

Other forms: Also available on the internet via WWW in PDF format.

Ctrct/Grnt/Proj/Task: Arthritis Research Campaign. EC069 1000

Related publications: Companion volume: RAND/TR-176-ARC.

Subject: Arthritis--Research--United States.

Subject: Research--Evaluation--Methodology.

Subject: Arthritis.

Personal name: Wooding, Steven.

Personal name: Hanney, Stephen.

Personal name: Buxton, Martin.

Personal name: Grant, Jonathan.

Corporate name: RAND Europe.

Corporate name: Rand Corporation.

Corporate name: Health Economics Research Group.

Corporate name: Arthritis Research Campaign (Organization)

Research unit: RAND Europe

Distribution code: 3

Distribution date: 20040923

The returns from arthritis research

Volume 1: Approach, analysis and
recommendations

Steven Wooding, RAND Europe

Stephen Hanney, Health Economics Research Group, Brunel University

Martin Buxton, Health Economics Research Group, Brunel University

Jonathan Grant, RAND Europe

MG- 251- ARC

Prepared for the Arthritis Research Campaign

Approved for public release; distribution unlimited



Preface

This report, prepared for and funded by the Arthritis Research Campaign (arc), presents the results of an evaluation of 16 research grants awarded by arc in the early 1990s. The main objective was to develop a system for evaluating arthritis research, with a view to allowing arc to stimulate and manage the exploitation of research advances so that they translate into outcomes of practical benefit to people with arthritis.

The report is organised into two volumes. Volume 1 is an overall analysis of the payback from 16 research grants. It presents a framework that conceptualises the relationship between research inputs, process, output and outcomes. Using this framework, we catalogue a diverse range of research output and outcomes arising from these 16 grants and make a series of quantitative and qualitative assessments comparing, for example, payback from project grants *versus* programme grants. In conclusion, we make six observations.

1. There is a diversity of research payback.
2. The researcher is the key driver of research translation.
3. Short, focused project grants seem to provide value for money.
4. Intended and unintended flexibility in funding is used advantageously.
5. Referees' contributions to the peer-review process are of variable benefit.
6. The payback framework could be operationalised and embedded by arc.

The companion Volume 2 is a collection of the case studies. These case study reports all follow a similar format based on the conceptual model and provide a rich and detailed narrative on the payback of each research grant.

In addition to arc's trustees, senior management, staff, scientists, fundraisers, donors and people with arthritis, this report should be of interest to other research funding agencies and to evaluators who are concerned with measuring the impact of science.

The research was led by RAND Europe in collaboration with the Health Economics Research Group at Brunel University. In addition, we commissioned bibliometric support from City University's Department of Information Science. RAND Europe is an independent not-for-profit policy research organisation that serves the public interest by improving policymaking and informing public debate. Its clients are European governments, institutions, and firms with a need for rigorous, impartial,

multidisciplinary analysis of the hardest problems that they face. This report has been peer-reviewed in accordance with RAND's quality assurance standards (for more information, see <http://www.rand.org/about/standards/>) and therefore may be represented as a RAND Europe product.

For more information about RAND Europe, please contact:

Martijn van der Mandele

President

RAND Europe

Newtonweg 1

2333 CP Leiden

The Netherlands

Tel: +31 71 524 5151

Fax: +31 71 524 5191

Email: mandele@rand.org

Jonathan Grant

Director

RAND Europe Cambridge

Grafton House

64 Maids Causeway

Cambridge CB5 8DD, UK

Tel: +44 (0)1223 353329

Fax: +44 (0)1223 358845

Email: jgrant@rand.org

Contents

Preface	iii
Table of figures	ix
Table of tables	xi
Executive summary	xiii
Acknowledgements	xxi
CHAPTER 1 Introduction.....	1
1.1 The Arthritis Research Campaign	1
1.1.1 Burden of arthritic disease in the UK.....	2
1.1.2 Support for arthritis research	2
1.1.3 Strategic review – <i>Research into Practice</i>	3
1.2 Origins and objectives of evaluation	4
1.3 Research evaluation	4
1.4 Organisation of report	6
CHAPTER 2 Approach.....	9
2.1 Evaluation framework.....	9
2.1.1 Payback categories	12
2.1.2 Payback model	15
2.2 Case study selection	19
2.2.1 Generating a selection matrix	20
2.2.2 Populating the selection matrix	23
2.2.3 Selecting the individual grants	25
2.2.4 Approaching principal investigators	25
2.3 Data collection/methods	25
2.3.1 Document and literature review.....	25
2.3.2 Semi-structured key informant interviews.....	25
2.3.3 Bibliometrics indicators	26
2.3.4 Clearance and validation.....	28
2.4 Case study pilots.....	28
2.5 Cross-case analysis	28
2.5.1 Qualitative analysis	28
2.5.2 Quantitative analysis.....	29

CHAPTER 3 Results.....	33
3.1 An analysis of payback categories	33
3.1.1 Knowledge production	33
3.1.2 Research targeting and capacity building	34
3.1.3 Informing policy and product development.....	37
3.1.4 Health and health sector benefits.....	39
3.1.5 Broader economic benefits	39
3.2 Analysis of the payback model	40
3.2.1 Stage 0: topic/issue identification.....	40
3.2.2 Interface A: project specification and selection (peer review)	40
3.2.3 Stage 1: inputs to research.....	46
3.2.4 Stage 2: research processes.....	47
3.2.5 Stage 3: primary outputs from research	48
3.2.6 Interface B: dissemination.....	48
3.2.7 Stage 4: secondary outputs – policymaking and product development	48
3.2.8 Stage 5: adoption – by practitioners and public	48
3.2.9 Stage 6: final outcomes	49
3.3 Cross-case analysis.....	49
3.3.1 Payback profiles – scoring consistency	49
3.3.2 Possible confounding variables.....	50
3.3.3 Mode of funding.....	52
3.3.4 Type of research	54
3.3.5 Bibliometric impact of researcher	55
3.3.6 Knowledge production as a predictor of health and health sector benefit.....	58
Closing remarks	59
CHAPTER 4 Conclusions and recommendations	61
4.1 Overview.....	61
4.2 Study limitations and caveats	62
4.2.1 Linearity of the payback model	62
4.2.2 Use and generalisation of case studies.....	62
4.2.3 Selection of case studies	62
4.2.4 Attribution and analysis of payback	63
4.2.5 Elapsed timescale	63
4.3 Policy observations and recommendations	63
4.3.1 There is a diversity of research payback.....	64
4.3.2 Individuals translate research.....	64
4.3.3 Short focused project grants seem to provide value for money	68
4.3.4 Intended and unintended flexibility is used advantageously	69
4.3.5 Referees' contributions to the peer review process are of variable benefit.....	69
4.3.6 The payback framework could be operationalised and embedded by arc.....	70
4.4 Concluding comments	72
REFERENCES	75

APPENDICES	83
Appendix A: Ranking researchers for case study selection	85
Summary.....	85
Classifying by funding type.....	85
Classifying by area of work.....	86
Classifying by success	88
Background	88
Identifying the papers.....	89
Database coverage and researcher type.....	89
Dealing with homonyms.....	89
Identifying a core set of papers	90
Identifying the remaining papers: co-author inclusion.....	90
Scoring papers.....	95
Journal Impact Factor	96
Scaled JIF	96
Fractionated JIF.....	96
Examining the rankings	97
Different methods of scoring provide similar rankings	97
Variation in ranking is highest for mid-impact researchers	97
Closing remarks.....	100
Qualifications used for classification	100
Appendix B: Case study project descriptions	103

Table of figures

Figure 2.1: Genesis and development of the payback model	11
Figure 2.2: Payback model adapted for the arc study	17
Figure 2.3: The original selection matrix.....	22
Figure 2.4: The final selection matrix.....	23
Figure 2.5: Feedback provided to team members during case study scoring.....	30
Figure 2.6: Two groups of “payback profiles” generated with artificial data.....	31
Figure 3.1: Peer-review assessment of six project grants and two programme grants.....	46
Figure 3.2: Payback profiles grouped by potential confounding factors	51
Figure 3.3: Payback profiles by mode of funding.....	54
Figure 3.4: Payback profiles by area of funding	57
Figure 3.5: Payback profiles by bibliometric impact of researcher	57
Figure 3.6: Payback profiles by level of knowledge production.....	58
Figure 4.1: Payback pathways for arc-funded research	66
Figure A.1: Methods used to classify arc researchers into basic scientists, clinicians or allied health professionals	86
Figure A.2: Flowchart of CAI method.....	91
Figure A.3: Percentage of papers found for each PI by two different techniques.....	92
Figure A.4: Comparison of the effectiveness of successive rounds of CAI	93
Figure A.5: A worked example of CAI	94
Figure A.6: Number of PIs and the fraction of their papers missed by ARC and ARTHR filters	95
Figure A.7: Equations used to calculate PI publication scores	96
Figure A.8: Comparisons between ranking when score is determined by summation of JIF, fractionated JIF or scaled JIF	98–99
Figure A.9: Relationship between ranking of researchers and total bibliometric score (as sum of JIF) for basic scientist PIs	99

Table of tables

Table 1.1: The evaluators' toolkit.....	7
Table 3.1: The outputs and outcomes of arc funded research – payback from 16 case studies	35–36
Table 3.2: Comparison of bibliometric summary data.....	37
Table 3.3: Payback model populated for 16 case studies	42–45
Table 3.4: Degree of disagreement in scoring case studies.....	50
Table 3.5: Qualitative analysis of payback by mode of funding	53
Table 3.6: Qualitative analysis of payback by type of research.....	56
Table 3.7: Qualitative analysis of payback by track record of researcher.....	60
Table 4.1: Summary of research paybacks	65
Table A.1: Selection matrix.....	85
Table A.2: Outline of method used to generate grant shortlists and issues raised by each stage.....	87

Executive summary

The mission of the Arthritis Research Campaign (arc) is to improve the lives of people with arthritis. arc aims to achieve this mission by raising funds to support medical research into the cause, treatment and cure of arthritic conditions. arc is the UK's fourth largest medical research charity, investing £22 million a year in research into arthritis. Currently, clinical and basic scientific research is supported through approximately 400 project grants, programme grants and fellowships in universities and medical schools throughout the UK. arc also provides core funding for its two major research institutes, the Kennedy Institute of Rheumatology in west London, and the Epidemiology Research Unit (arc ERU) at the University of Manchester.

To mark its 65th anniversary in 2002 arc decided to undertake a strategic review that resulted in the publication of a five-year strategic plan, *Research into Practice*. The review was informed by consultations with arc's stakeholders – including trustees, staff, scientists, volunteer fundraisers, donors and people who have arthritis – and concluded that “there seems to be a gap between the aspirations of people affected by arthritis and the ability of science and academia to meet those aspirations”. In order to bridge this gap, arc decided to “instigate a system of rigorous retrospective evaluation on work which has already been completed, with a view to identifying opportunities for development”. To inform this commitment, arc commissioned this study to:

- review and document the long-term outcomes of arc research grants awarded in the early 1990s;
- identify the factors associated with the successful translation of research;
- illustrate the strengths and weaknesses of different modes of funding; and
- identify “good news stories” that arc could use in its public engagement and fundraising activities.

The purpose of this volume is to report on the approach, results, conclusions and recommendations arising from an in-depth evaluation of 16 research grants funded by arc in the early 1990s. It is supported by a second volume, *The returns of arthritis research. Volume 2: Case studies*, that describes the output and outcomes from the grants. In this executive summary we set out what we did, reporting our key conclusions and their implications for policy.

Evaluation purpose and approach

This evaluation is intended to improve understanding of how research is translated from “bench to bedside”. It examines the historical development of 16 case study research

grants, and assesses the extent to which different types of funding support might prevent or promote the successful translation of research.

To conduct this inquiry, the research team created a framework that breaks down the process by which research translates into practice. The framework had two elements. The first element is the five payback categories (summarised in Box S.1). The second element is the payback model (illustrated in Figure 2.2, Chapter 2 and summarised in Box S.2). The payback categories and model were adapted from the Buxton and Hanney Payback Framework following interviews with a series of key informants.

- Knowledge production
- Research targeting and capacity building
- Informing policy and product development
- Health and health sector benefits
- Wider economic benefit

Box S.1: Payback categories

Stage 0:	Topic/issue identification
Interface A:	Project specification and selection (peer review)
Stage 1:	Inputs to research
Stage 2:	Research process
Stage 3:	Primary outputs from research
Interface B:	Dissemination
Stage 4:	Secondary outputs
Stage 5:	Adoption
Stage 6:	Final outcomes

Box S.2: Summary of payback model

Guided by this framework, we conducted case studies of 16 research grants. The case studies were selected from 556 possible grants awarded by arc between 1990 and 1994. In order to allow us to compare the effect of the mode of research support, the type of research and the bibliometric impact of the principal investigators (PIs), we constructed a selection matrix. With the help of the Development Committee, we chose six project grants, three programme grants, three fellowships and four institute grants for evaluation. Our collection of grants contained six basic grants, eight clinical grants and two allied health professional (AHP) grants (classified according to the qualifications of the PIs), with nine "high" impact PIs and seven "mid" impact PIs.¹ With 16 case studies we could

¹ As explained in Appendix A, impact was measured using a range of different bibliometric indicators; "high" impact was the top decile of PIs based on those indicators, and "mid" was the 45–55 percentile.

not expect them to be representative of all arc grants in a statistical sense; however, by using a selection matrix we aimed to produce a set of case studies that mirrored the diversity of arc funding in key dimensions and hence from which could be carefully generalised.

Using the information collected from document and literature reviews, semi-structured key informant interviews and bibliometric analysis, each of the 16 cases was written up as a narrative organised according to the structure provided by the payback framework (Box S.2). Using a common structure facilitates comparative analysis, allowing us, for example, to identify the factors associated with the successful translation of research. We employed two approaches to our cross-case analysis. The first was based on a qualitative assessment of the case studies based on a discussion within the project team of the key observations made by each member of the team. The second involved a novel method of scoring the case studies on the five payback categories.

Conclusions and implications for policy

The study reached six main conclusions, which we discuss below. However, there are several limitations to our approach, the key issues being:

- whether it is reasonable to use a largely linear framework to structure analysis of the scientific process;
- whether the use of, and generalisation from, case studies, is appropriate;
- biases in the process of selecting our case study grants;
- how to determine whether a specific outcome can be attributed to a particular grant or investigator;
- how to pick a suitable time window for the start of study: a compromise between allowing outcomes to come to fruition and ensuring that records are available and investigators' recall are suitably detailed.

Each of these issues is discussed in more detail in Chapter 4 (section 4.2). By discussing them we do not wish to undermine our conclusions, but to illustrate some of the challenges of evaluating research.

There is a diversity of research payback

There is strong evidence from our analysis that there is a considerable range of research paybacks and that these would not have been identified without the structured, case study approach employed in this evaluation. The highlights of these paybacks are listed in Table S.1.

Table S.1: Summary of research paybacks

Payback category	Payback	Example
Knowledge production	Peer-reviewed publications in the serial literature	302 papers receiving a total of 975 citations per year attributable to case studies
Research targeting and research capacity	Postgraduate research training Subsequent career development of PIs and research assistants The transfer of technical know-how Informing future research studies	28 PhD/MDs from work on the case studies Development of technological know-how in genetic mapping Informed >£2 million Medical Research Council (MRC) randomised controlled trial Use of biologicals as therapeutic targets
Informing policy and product development	Informing recommendations in clinical guidelines and other policy advice Informed development of clinical tests	Recommendation in Royal College of Obstetricians and Gynaecologists (RCOG) guideline on the use of aspirin and heparin for women with antiphospholipid syndrome (APS) Recommendation in Industrial Injury Advisory Council (IIAC) assessment for hip osteoarthritis (hip OA) in farmers to be a prescribed disease Clinical test for a rare type of systemic lupus erythematosus (SLE) and chondrodysplasia type Schmidt
Health and health sector benefits	Improving the quality of life for people with rheumatoid arthritis (RA) Reducing the likelihood of recurrent miscarriages for women with APS	Hundreds of thousands of patients treated with anti-TNF of whom 70% experience a significant improvement in health Use of aspirin and heparin for women with APS increases live birth rate by 40% compared to the use of aspirin alone and by 60% compared to no treatment at all.
Wider economic benefits	Unquantified economic returns resulting from a reduction in days off work and sales of licensed drugs	

Individuals translate research

There is good evidence from our 16 case studies that when translation of research into developments of practical value to patients occurs it is largely due to the conviction, effort

and personal networks of a particular investigator, and is not associated with the type or mode of the funding stream or the bibliometric impact of the investigator. This complements previous studies that have shown that encouraging partnership between researchers, practitioners, policymakers and industrialists promotes successful translation.

Therefore, we propose that **arc** introduces two new types of award. “Translation awards” would be topic-focused and directly linked to the translation of a previous piece of **arc**-funded research. “Partnership awards” would be people-focused and provide resources to **arc**-funded researchers to develop networks with potential users of research. This could include supporting secondments to or from commercial and non-commercial laboratories, and participation in policymaking networks. Criteria for translation and partnership would focus on the potential return or payback from translation, the stated route or plan of translation, relevance to **arc**’s strategic aims and, in the case of translation awards, evidence of existing networks.

Short focused projects grants seem to provide value for money

There is good evidence from our analysis that the payback arising from projects grants is similar to that arising from the other modes of funding. Given that the median value of a project grant is £90,000 (compared to £250,000 for fellowships, £480,000 for programmes and £450,000 for institutes) this indicates that they provide significant value for money. Of all the observations that we have made from our analysis, this was the most unexpected and surprising and illustrates the importance of maintaining a funding mechanism for short-term, focused research of this nature.

Intended or unintended flexibility in funding is used advantageously

There is some evidence from our case studies that investigators successfully exploit flexibility in the scientific and administrative management of grants. In none of the case studies was there any evidence that this flexibility had a negative effect on the scientific outputs and outcomes of the research, and in some cases there were indications that such flexibility was used advantageously. This observation therefore supports the continuation of **arc**’s current policy of flexibility in funding.

Referees’ contributions to the peer-review process are of variable benefit

There is some evidence from analysis of successful applications that referees’ contributions to review panels do not add significant scientific value to the reviewed proposals. However, it is worth noting that the primary purpose of the review process was to select suitable applications for funding, rather than to improve those successful applications. For nine of the case study proposals, even where the referees’ comments were fed back to the PI, they had little or no impact. For four of the case studies, the peer-review process did have an impact on the design of the study. For a further two cases (which had the highest payback), if the referees’ comments had been taken at face value and not overruled by the assessing panel, the proposed research would not have been funded.

The payback framework could be operationalised and embedded by **arc**

There is good evidence from this study that the payback framework adapted for **arc** works and, given the appropriate management information, could be operationalised prospectively to stimulate and manage the returns from **arc** research. The payback framework proved to be effective in capturing the diverse range of research outputs and outcomes, and in identifying cases where research had been translated to benefit people with arthritis. If applied prospectively, the framework could be used to inform the granting of the recommended translation and partnership awards. (In Chapter 4 we

describe how **arc** could operationalise and embed the payback framework, and identify a number of issues that would need to be resolved prior to implementation.)

Recommendations

On the basis of these conclusions we make six recommendations, which are intended to help **arc** develop a system to ensure the successful translation of the research that it funds. These are outlined below, along with the aim and context of each recommendation.

Recommendation	Context	Aim
1. arc should survey all forms of payback when monitoring and evaluating the returns from arthritis research.	There is strong evidence from our case studies that all types of grant produce a range of research outputs and outcomes, beyond the usually assessed publications in the peer-reviewed literature.	To ensure that the returns from arc -funded research are fully recorded and recognised.
2. arc should selectively support investigators in translating their research. This might include: translation awards to promote the successful transfer of knowledge with potential health benefit; interaction awards to develop productive relationships between researchers and policymakers or industry. These awards could be made in both reactive and directed modes.	There is good evidence from our case studies that when translation occurs, it is largely down to the individuals' conviction, effort and personal networks, although individuals currently have little or no support for these activities.	To recognise the importance of personal networks in the translation of research, and to ensure that translation opportunities are resourced fully and realised.
3. arc should continue to support short focused project grants as part of its funding portfolio.	There is good evidence from our case studies that project grants provide value for money when compared to programme grants, fellowships and institutes.	To acknowledge the importance of project grants in funding research.
4. arc should maintain its flexible approach to the funding and administration of research grants. In addition we suggest that arc considers the costs and benefits of fixed budget funding.	There is some evidence from our case studies that investigators successfully exploit flexibility in the scientific and administrative management of grants.	To confirm that arc should maintain its current policy of being flexible in the award and administration of grants.
5. arc should review its peer-review processes to maximise their efficiency and effectiveness.	There is some evidence from our case studies that for successful applications referees' contributions to review panels are of variable benefit.	To challenge arc into assessing the costs and benefits of its peer-review system, with a view to improving its value to applicants.
6. arc should consider developing systems for the ongoing and prospective monitoring and evaluation of its funded research.	There is good evidence from this study that the payback framework developed for arc works and, given the appropriate management information, could be operationalised to prospectively monitor the returns of arc -funded research.	To develop an approach whereby arc will be in a position to "stimulate and manage the exploitation of research ... into outcomes of practical benefit to people with arthritis".

Acknowledgements

The Arthritis Research Campaign commissioned this study. Jonathan Grant and Steven Wooding from RAND Europe, in collaboration with Martin Buxton and Stephen Hanney from the Health Economics Research Group at Brunel University, wrote the successful proposal. The subsequent project team involved seven members. Steven Wooding managed the project, with the support and guidance of Jonathan Grant.

Steven Wooding, Steve Hanney, Martin Buxton and Jonathan Grant devised the methodological approach and analysis, as report in this volume. The conclusions that emerged in the final chapter of Volume 1 resulted from a workshop in which all members of the project team participated. Different researchers conducted the case studies reported in this volume, as described in the introductory chapter of Volume 2 of the report. In summary, the case study authors were: Silvia Anton, Jonathan Grant, Stephen Hanney, Stijn Hoorens, Abigail Lierens, Miriam Shergold and Steven Wooding.

As is the nature of a project of this scale and complexity, the project team would like to acknowledge the invaluable support provided by a number of groups and individuals. First, Grant Lewison, Isla Rippon and Kate Wilcox-Jay, all from City University's Department of Information Science, who supplied us with the bibliometric data and analysis that underpin the study. Second, the important role of the Arthritis Research Campaign's Development Committee who acted – in the true meaning of the word – as a steering group and provided excellent guidance during all stages of the study. The Development Committee was made up of the following members: Patrick Sissons, Ann Raven, Fergus Logan, Madeleine Devey, Michael Patnick, Mary Collins, Matthew Brown, Mike Hurley and Tony Woolf. Third, in addition to their participation in the Development Committee, we would also like to acknowledge the role of Arthritis Research Campaign staff, most notably Fergus Logan, Madeleine Devey and Michael Patnick, for initiating the study and for answering our numerous queries. We would also like to thank Charlene Rohr and Dr Cyril Frank for providing very helpful comments and suggestions in reviewing the report, and Lisa Cordaro for copy editing the report.

We would like to reserve our final acknowledgements to all those scientists – 46 in total – who were willing and able to act as our experimental subjects for this study. This involved either being the subject of a case study or being interviewed as a stakeholder of a case study. As is the nature of this type of evaluation it is inappropriate to name these individuals, but without their support and commitment this study would not have been possible.

Measuring the returns from research is of growing importance to research-funding organisations. This is because they are under increasing pressure to demonstrate the benefits arising from their expenditure which (in the case of government agencies) is directly funded from the taxpayer, or (in the case of charitable organisations) indirectly supported by the taxpayer through fiscal benefits and directly supported by philanthropic donations. In addition to demonstrating accountability and good research governance, research-funding organisations need a robust evidence base to inform strategic decisions on how to fund research.

The purpose of this volume is to report on the approach, results, conclusions and recommendations arising from an evaluation of a sample of research grants funded by the Arthritis Research Campaign (arc) in the early 1990s. It is supported by a second volume, *The returns of arthritis research. Volume 2: Case studies* (Wooding and others 2004), which describes a collection of the case studies compiled as part of the evaluation.

The remainder of this chapter describes the origins of the study, provides some background information on arc, reviews the various methods that are used in research evaluations and sets out how this volume is organised.

1.1 The Arthritis Research Campaign

arc was founded in 1936 and raises funds to:

- promote medical research into the cause, treatment and cure of arthritic conditions;
- educate medical students, doctors and AHPs about arthritis;
- provide information to people affected by arthritis and to the general public.

The mission of arc is to improve the lives of people with arthritis.

Currently, clinical and basic scientific research into arthritis is supported through approximately 400 project grants, programme grants and fellowships in universities and medical schools throughout the UK. Both clinicians and scientists are encouraged to undertake rheumatology or musculoskeletal research through arc's fellowship and studentship schemes.

Support is provided for clinical rheumatology units, particularly in medical schools, to enable them to integrate their clinical, research and teaching activities. arc provides core funding for its two major research institutes, the Kennedy Institute of Rheumatology in

west London and the Epidemiology Research Unit (arc ERU) at the University of Manchester.

arc is also dedicated to the enhancement of rheumatology teaching for medical undergraduates and the development of academic rheumatology. To achieve this, it has established 10 chairs in rheumatology and, in addition, has provided clinical academic posts at lecturer or senior lecturer level in 16 medical schools in the UK.

1.1.1 Burden of arthritic disease in the UK

There is no national system for monitoring the musculoskeletal health of the population and the extent of arthritis and related conditions. For this reason, in May 2002 arc published two separate studies under its "Arthritis: the Big Picture" campaign: one from MORI and the other from the arc ERU, to find out how many people are affected by arthritis (ARC 2002a). More recently, Arthritis Care commissioned an omnibus survey to establish the scale of the incidence of arthritis in the UK (Arthritis Care 2002).

These three studies indicate that between 7 and 13 million people in the UK are affected by arthritis (see Box 1.1). The discrepancy in the figures may relate to confusion over what constitutes arthritis, joint pain or musculoskeletal pain as well as reflecting the different ways in which the estimates were calculated. Nevertheless, these three independent studies indicate that around one in six of the UK population has arthritis. The MORI study showed also that arthritis and joint pain is more prevalent in women, the old, poor and uneducated, while the arc ERU study estimated that 206 million working days were lost due to arthritis (equivalent to £18 billion). The estimated direct cost to health and social services was £5.5 billion, while prescription costs added up to £341 million, and hip and knee replacements costs were £405 million.

1.1.2 Support for arthritis research

arc is the UK's fourth largest medical research charity,² investing some £22 million a year in research into arthritis (ARC 2004). It is notoriously difficult to estimate how much money is spent on a specific research field. One approach is to analyse research publications that acknowledge funding sources.³ A 1999 analysis of UK biomedical research papers concluded that arc is "in a dominant position within the UK in the arthritis subfield"; a quarter of papers that acknowledge funding sources mention arc, compared to approximately 13% for the Medical Research Council (MRC) and approximately 7% for the Wellcome Trust (Lewison and Devey 1999).

² The others are the Wellcome Trust, Cancer Research UK and the British Heart Foundation.

³ It should be noted that not all authors acknowledge funding, and typically 40% of all papers do not have a funding body acknowledgement.

How many people have arthritis?	
arc ERU	7 million
MORI survey	13 million
Arthritis Care survey	9 million
Who suffers from arthritis or joint pain?	
34% of women, compared to 23% of men	
52% of people aged 55 or over, compared to 18% aged under 55	
What are the costs of arthritis?	
£18 billion in lost working days	
£5.5 billion direct costs for health and social services	
£341 million for prescription costs	
£405 million for hip and knee replacements	
Managing the effects of arthritis?	
81% of people with osteoarthritis experience constant pain	
54% of adults with arthritis do not take drugs	
54% of adults with arthritis have taken complementary medicine	
66% of people are satisfied with the care that they receive from their GP	

[SOURCES: ARC 2002a; Arthritis Care 2002]

Box 1.1: Arthritis – facts and figures

1.1.3 Strategic review – *Research into Practice*

In 2002, to mark its 65th anniversary arc decided to undertake a strategic review of its activities and impacts. This resulted in the publication of a five-year strategic plan, *Research into Practice* (ARC 2002b). The review was informed by consultations with arc's stakeholders – including trustees, staff, scientists, volunteer fundraisers, donors and people who have arthritis – and concluded that “there seems to be a gap between the aspirations of people affected by arthritis and the ability of science and academia to meet those aspirations”. As a result the strategy document goes on to state that the “arc now plans to establish mechanisms to bridge this gap, and to stimulate and manage the exploitation of research and educational advances so that they translate into outcomes of practical benefit to people with arthritis”.

In recognition of the consequences of this strategic shift from research outputs to research outcome, arc needs to “instigate a system of rigorous retrospective evaluation on work which has already been completed, with a view to identifying opportunities for development” (ARC 2002b). To oversee this process, arc appointed a development committee (see Box 1.2 for its remit) which commissioned the study reported here.

- Evaluate the success or otherwise of arc-funded work
- Identify future potential
- Recommend initiatives to build on individual pieces of work
- Encourage and coordinate collaboration in emerging areas of interest and progress
- Establish priority areas

Box 1.2: Remit of the arc Development Committee

1.2 Origins and objectives of evaluation

arc commissioned the evaluation of arc-funded research in April 2003 to support the Development Committee in its deliberations. The objectives of the study are set out in Box 1.3 and were agreed between the project team and the Chief Executive of arc, Fergus Logan. The evaluation began in May and was scheduled to last 15 months. Throughout the evaluation the project team were supported by the Development Committee and arc's senior management.

- Review and document the long-term outcomes of arc research grants awarded in the early 1990s.
- Identify factors associated with the translation of research, and begin to develop "early success indicators" that can facilitate the translation of research into practice and can be used to help arc's Development Committee to fulfil its remit.
- Illustrate the strengths and weaknesses of different modes of research funding, which could inform current practice.
- Identify "good news stories" and vignettes of the research process that the arc could use in its public engagement and fundraising activities.

Box 1.3: Evaluation objectives

1.3 Research evaluation

As noted above, there is a growing need for funding agencies to evaluate the impact of (biomedical and health) research in order to inform their own research strategy, justify public expenditure on research and provide an enhanced impetus for research translations activities (Council on Health Research for Development (COHRED) 2000; Grant and others 2004; National Institutes of Health 2000; National Audit Office (NAO) 2003a; Smith 2001).

However, measuring the performance and results of research is challenging and complex. Many research impacts are not easily quantifiable and it is difficult to attribute a policy or clinical impact to a particular research result. In the US the Government Performance and Results Act of 1993 required federal agencies – including those that fund research – to set strategic goals and to use performance measures for management and budgeting. In Canada, the report of the Auditor General of Canada (1994) concluded that departments

and agencies should establish the mechanisms and practices that they need to demonstrate the results of their science and technology activities and to ensure that their resources are allocated effectively. In the UK, the NAO recently concluded that government departments "have no systematic mechanisms for measuring the overall impact of their research effort" (NAO 2003a).

As a result of the US Government Performance and Results Act of 1993, in 1998 the Committee on Science, Engineering and Public Policy (COSEPUP) held a series of workshops to generate ideas on how to evaluate research. These workshops identified and assessed six methods of evaluating research. These methods, along with others subsequently identified or developed by the authors, are reviewed in Table 1.1 (below).

At a similar time, the UK Department of Health commissioned a literature review on research evaluation methodologies (Cronin and Normand 1998). This review identified three issues inherent in evaluating research programmes. First, is the relationship between an evaluation and the strategic framework of the funding organisation. For example, if the mission of *arc* was to "generate new knowledge", then a bibliometric assessment of research publications may be entirely appropriate. However, as *arc*'s mission is to improve the lives of people with arthritis, then it is important to understand how new knowledge is (or is not) translated into effective clinical practice or preventive medicine.

Second, an evaluation needs to consider the type of research that is being undertaken. For example, bibliometrics provides an indicator of knowledge production that is more robust for basic research (which is published typically in the peer-reviewed literature), than applied research (which may be published in the more targeted but relevant "grey literature"). Similarly, the Buxton and Hanney Payback Model has shown that it is possible to evaluate the payback of applied health services research on policy and practice (Buxton and Hanney 1996), but this approach has been tested only recently on early clinical and basic research (Buxton and Schneider 1999; Hanney, Frame and others 2003).

The final issue is that the tendency to use single indicators is misleading, and that all research evaluation criteria need a qualitative and quantitative information base. For example, an analysis of the payback of Alberti's early clinical research in diabetes illustrated that some papers with few citations were considered to be of significance by his peers and other stakeholders, as elicited from interviews (Hanney, Frame and others 2003). More generally, that study showed that bibliometric data could inform the qualitative approach and provide indicators against which to correlate or challenge accounts from interviews.

To meet the evaluation objectives that were set by *arc* (shown in Box 1.3 and as discussed in Chapter 2), we developed an evaluation methodology based on a number of streams of our previous studies. This included the adaption and refinement of the Buxton and Hanney Payback Model and its Payback Categories (Buxton and Hanney 1996), the application of bibliometric techniques to that model (Grant 1999; Grant and others 2000), and the combination of qualitative and quantitative information and the prospective tracing of research learned from the 'Alberti study' (Hanney, Frame and others 2003). In addition (and as discussed in Chapter 2, section 2.2), we used an innovative way of selecting researchers for case study selection and developed a new way of summarising evaluation criteria in 'Payback Profiles'. As a result, not only does the work reported in this volume evaluate the long-term outputs and outcomes of *arc*-funded

research, but it also makes a number of contributions to the field of research evaluation methodologies.

1.4 Organisation of report

The remainder of this volume is organised in a classic structure. Chapter 2 sets out our evaluation approach. It describes the payback model and categories developed for arc, the selection of 16 case studies, the methods used in compiling the case studies, including various bibliometric indicators and how we undertook the cross-case analysis. The results of the evaluation are provided in Chapter 3. This includes the cataloguing of research outputs and outcomes from the case studies and an assessment of the commonalities and differences between different funding modes (that is, for example, project grants versus programme grants) and funding types (that is, for example, clinical research versus basic research). Chapter 4 elucidates our conclusions and makes a number of recommendations that are pertinent to the arc's research policy and its evaluation approach.

As noted at the beginning of this chapter, it should be stressed that this volume should be read alongside the case study report, *The returns from arthritis research. Volume 2: Case studies* (Wooding and others 2004).

Table 1.1: The evaluators' toolkit

Method	What is it	Advantages	Disadvantages
Bibliometric analysis (eg Wellcome Trust 2001)	Assessment of research publications Includes assessment of quantity, quality, collaboration	Quantitative Useful to see large-scale trends Repeatable analysis possible	Estimates of "quality" may not be reliable Difficult to compare across fields May be skewed by biases in data
Economic rate of return (eg Access Economics 2003)	Assess rate of return on research investment, eg how many £s returned for every £ spent	Quantitative Powerful for political lobbying	Focus on financial value of social or health benefits Relies on many assumptions that may be controversial and unreliable Unfamiliar to research community
Peer review (eg Woodring and Grant 2003)	Qualitative assessment by peers similar to that used for grant proposals or research publications	Well-understood and accepted by researchers Based on specialised knowledge Flexible and sensitive to discipline	Perceived biases Time consuming for experts Lack of transparency
Case studies (eg Hanney and others 2003)	In-depth examination of research projects Provides "narrative" of research process and outcomes	Provides in-depth understanding Informs reforms in the system Illustrates all types of research benefits	Hard to compare Single study may not be representative
Retrospective analysis (eg Grant and others 2004)	Historical assessment of the research process and outcome	Useful for identifying linkages between funding programmes and innovations over time	Issues of recall bias Relies on the availability of archives Not useful for short-term as lag between research and outcomes may be many years
Logic models (eg Buxton and Hanney 1996)	A series of "if ... then" statements which provides a picture of how a (research) programme works	Value in creating, validating and modifying model Provides structure to evaluation Allows comparisons between projects and programmes Easy to use	Tends to simplify research process Inflexible
Benchmarking (eg NAO 2003b)	Comparing across different countries, organisations or programmes	Generates ideas Allows for learning from other systems Allows identification of different practice	Difficult to interpret Affected by context

(SOURCES: COSEPIP 1999; Wellcome Trust 2001; feedback and questions from various seminars and workshops where variants of this table have been presented)

The Arthritis Research Campaign (arc) wants to develop a system of evaluating the outcomes of the research that it funds, with the view to identifying opportunities for future development. This report deals with a sample of 16 research grants funded in the early 1990s and evaluates the research outputs and outcomes arising from those grants over a 10-year period. The approach applied and developed for this evaluation is described in this chapter. We begin by explaining the evaluation framework that was adopted for the study and how we selected the 16 research grants from a possible 556 grants. We then detail our data collection methods – including the use of document reviews, key informant interviews and bibliometrics – and how we analysed the case studies in order to draw out our conclusions and recommendations for arc.

As discussed in the final chapter, it is envisaged that the approach described below could be further developed in part to allow arc to operationalise the continuous evaluation of its research, thus allowing it to encourage proactively the exploitation of research into outcomes that benefit people with arthritis.

2.1 **Evaluation framework**

The evaluation framework is made up of two components. The first component is the definition of evaluation criteria for the outputs and outcomes of research. It should be stressed that evaluation criteria can be quantitative (eg the number of research publications) or qualitative (eg a description of career progression following the awarding of a research grant). The second component is a logic model of the research process. Logic models are widely used in evaluation methodology to understand input–process–output relationships and to break down research programmes into their constituent parts (WK Kellogg Foundation 2001).

For the present study we adapted the evaluation criteria and logic model developed by the Health Economics Research Group at Brunel University to assess the “payback” of health services and biomedical research. This work was commissioned originally by the Department of Health in 1993 to evaluate the health service research that it supported. Subsequently the payback framework has gone through a number of iterations and applications. The first phase of the work, described in Buxton and Hanney (1994, 1996) and Buxton and others (1994), consisted of:

a categorisation of payback under five headings: knowledge, research benefits, political and administrative benefits, health sector benefits and broader economic benefits (as illustrated in Box 2.1);

the development of a nine-step model consisting of seven stages and two interfaces showing how, and at what stages, categories of payback could be assessed (as illustrated in Figure 2.1); and

the testing of the categorisation and modelling in eight case studies.

The second phase of the study (Buxton and Hanney 1997; Hanney and Buxton 1997), confirmed that the multidimensional categorisation of payback, as originally presented under the five headings listed above, was robust, although two additional subcategories were added (these are "B(iv)" and "D(v)" in Box 2.1). Similarly, in reviewing a further 10 case studies, it was shown that the nine-step model was valid, but the issue of whether the scientific endeavour can be modelled as a linear process and the importance of political and professional environment were raised. This led to further refinement of the payback model (as illustrated in Figure 2.1) with a final version presented in Hanney, Gonzalez-Block and others (2003).

A.	Knowledge
B.	Benefits to future research and research use:
	i. better targeting of future research;
	ii. development of research skills, personnel and overall research capacity;
	iii. critical capability to utilise appropriately existing research, including that from overseas;
	iv. staff development and educational benefits.
C.	Political and administrative benefits:
	i. improved information bases on which to take political and executive decisions;
	ii. other political benefits from undertaking research.
D.	Health sector benefits:
	i. cost reduction in the delivery of existing services;
	ii. qualitative improvements in the process of service delivery;
	iii. increased effectiveness of services, eg increased health;
	iv. equity, eg improved allocation of resources at an area level, better targeting and accessibility;
	v. revenues gained from intellectual property rights.
E.	Broader economic benefits:
	i. wider economic benefits from commercial exploitation of innovations arising from R&D;
	ii. economic benefits from a healthy workforce and reduction in working days lost.

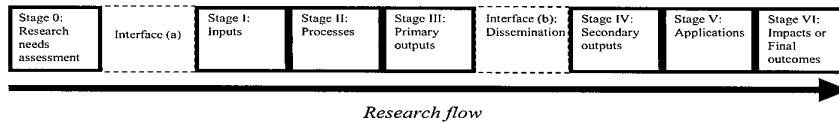
(SOURCE: Buxton and Hanney 1994, 1996, 1997)

Box 2.1: Payback categories

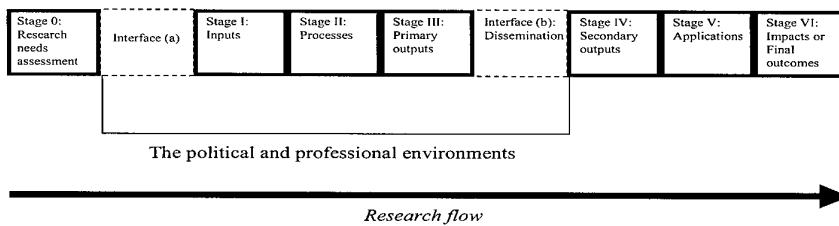
From this basis, the payback framework has been applied to a number of different contexts. For example, Buxton and Schneider (1999) explored the possible application of this to a Canadian research organisation that funded basic biomedical and early clinical studies alongside health services research. Buxton and others (2000), developed a postal questionnaire based on the payback categorisation and model to be applied to the (then) National Health Service's North Thames Research and Development Programme. This approach was developed further in Croxson and others (2001), which examined how a

system of monitoring could be developed. More recently, the model has informed an analysis of health research systems on behalf of the World Health Organisation (Hanney, Gonzalez-Block and others 2003; Pang and others 2003).

Basic model



The political and professional environments



Forward leaps and feedback loops

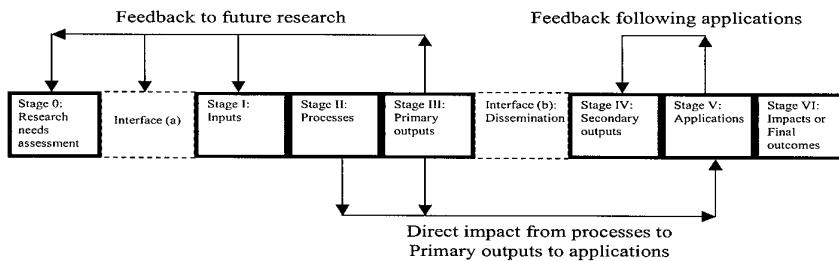


Figure 2.1: Genesis and development of the payback model

The lesson from the application of the payback framework is that it may need to be adapted in various ways in order to meet the particular circumstances of the research funder. Accordingly, at the outset of the study we undertook five interviews with key individuals within arc and the associated research field. The purpose of these interviews was to test and validate the payback framework in the context of arc-funded research. As a result of the interviews we reviewed and revised the payback framework and sought final agreement and sign-off from the Development Committee. The main changes involved the generalisation of the names of some of the stages to make them more obviously applicable to basic science and response mode funding, for example, "Stage 0: research needs assessment" was generalised into "Stage 0: topic/issue identification" (compare Figures 2.1 and 2.2).

2.1.1 Payback categories

The multidimensional category of payback provides the evaluation criteria for the outputs and outcomes from arc funding. Although a number of specific arthritis examples were cited, in our interviews no new categories or sub-categories were identified in addition to those listed in Box 2.1. The interviews and subsequent review of the model led us to generalise the titles of the categories to apply outside the area of health services research where they had been developed initially. Below each category is considered in turn, with various sub-categories explored and measures described.

Knowledge production

The knowledge produced by research is the first output and is contained in various publications and patent applications. The US National Science Board makes an annual assessment of national performance by publishing counts of scientific papers and patents in its "National Indicators" series (National Science Board 1996). Similarly, in the UK, the Research Assessment Exercise evaluation of university departments includes the submission of scientific papers as part of assessment procedures (Wooding and Grant 2003). More recently, the Office of Science and Technology (OST) has used bibliometric data for public sector agreement target metrics for the UK science base (Office for Science and Technology 2003). These types of bibliometric analyses have attracted their critics (Seglen 1997), not least because they have been used in isolation of other methodologies and have failed to use multiple indicators in the assessment of research (Martin 1996). At its most fundamental level, bibliometrics should be used to generate hypotheses rather than to provide conclusive evidence on a particular policy or intervention. In this study we have used a number of bibliometric indicators to inform our case studies (and the wider contextual analysis of the arthritis field), but these indicators are complemented by other information collected from the document and literature reviews and the key informant interviews.

In addition to counting the number of publications, their quality and impact can be assessed in various ways. Traditionally, the quality of knowledge production has been assessed by peer review, but various other methods can be applied. Often, papers that are accompanied by an editorial are seen as being of particular significance. For those studies that are included in a systematic review there are now formal quality assessment techniques (Grimshaw and Eccles 1998). Citation analysis can be applied to assess the impact that the specific article is having within the research community (Dawson and others 1998; Wellcome Trust 2001). Previous experience suggests that knowledge production will be particularly important for basic research, and certainly papers in basic research journals tend to be cited more frequently than those in clinical journals (Lewison and Dawson 1998).

A journal's "impact factor" is based on the average number of times that an article in the journal is cited; this can provide a shorthand version of citation analysis by giving some indication of the importance of the journal in which an article appears. The use of impact factors in analysis of biomedical research has been criticised (Lewison and Dawson 1998), but provided that care is taken (Garfield 1996), it has been shown to be of some value (Lewison and Devey 1999; Wellcome Trust 2001).

Particularly when considering research that might be aimed at potential users outside the research community, it is often desirable to use a range of publication outlets, including those journals with the highest readership among the groups at whom the research is targeted. In some fields these may well be journals that do not have an impact factor but that are, nevertheless, significant as vehicles for dissemination of the knowledge produced (Hanney, Soper and others 2003; Jones and others 2004; Royal Netherlands Academy of Arts and Science 2002).

Research targeting and capacity building

The better targeting of future research is frequently a key benefit from research, especially from research that is more basic and/or methodologically oriented. An indication of this comes from citation analysis. The targeting can be of both the research conducted by others and the original researcher(s). Where follow-on research (especially by members of the original research team) is clearly associated with the original research, it can be useful to obtain information on the source and amount of such funding. As is developed in the paragraph below, one of the key roles of a medical charity can be to fund research in its field that will help to open up questions or issues that will then attract further funding from general research funders such as the Medical Research Council (MRC) and the Wellcome Trust.

Research training can be provided both as a result of the employment of staff on research projects and programmes, and through explicit funding for research training and career development. One measure of research training which may appear crude but nevertheless has been used in previous studies, is the number and level of higher or research degrees resulting (either totally or in part) from the research funding (Mushkin 1979; Verhorn and others 1982). The career development of arthritis researchers goes much wider than specific training and is of considerable importance to arc, which aims to ensure that the pool of researchers in this field is as strong as possible. The reasoning is that this, in turn, should help to ensure that arthritis as a topic is able to gain an appropriate share of the research funding that is available from general medical research funders. Some of arc's funding schemes aim explicitly to provide career development, and for other researchers the receipt of a project grant from arc can be important in advancing their career in research.

Informing policy and product development

Research can be used to inform policymaking in a wide range of circumstances. Policymaking here is interpreted very broadly to cover not only government national policies, but also:

- polices made by managers at many levels within a health service;
- polices agreed at the national or local level by groups of health care practitioners in the form of clinical or local guidelines; and

policies developed by those responsible for training, education or inspection in various forms including training packages, curricula and audit and evaluative criteria (Hanney, Soper and others 2003).

Basic research is less likely than that from clinical researchers or allied health professionals (AHPs) to be used to inform policy.

On a similar level, although it involves very different processes, research can be used also to inform product development. Informing policies and product development are conceptually similar in that there generally has to be some subsequent adoption of the policy, or product, before the health and economic benefits can accrue.

Health benefits

These benefits might be viewed as the “real” payback or outcomes from health research. Greater clinical effectiveness resulting from research-informed drugs or procedures will lead to increased health. Various measures of health gain do exist, but in most cases for arthritis the emphasis is likely to be on those that assess reduction in pain or disability, or increased mobility. While the benefits from arthritis research will not be generally measured in terms of life years gained, in principle they might be captured by using Quality Adjusted Life Years (QALYs). In countries such as the UK, this is seen as a more appropriate approach than using Disability Adjusted Life Years (DALYs) (Fox-Rushby 2002). There have been recent attempts to put a monetary valuation on the reduction in mortality and morbidity as a result of health research (Access Economics 2003; Murphy and Topel 2003); however, there have been criticisms of these approaches (Buxton and others in press).

This category of benefits can be thought of as going wider than health gain, and some aspects can be seen as benefits to the health sector more generally. Cost savings in the provision of health care may result from research-informed changes in the organisation of services, or in the particular therapies that are delivered. It might be necessary to consider various issues here. These include whether potential savings have been realised in practice – either as cash savings or the release of resources for other valuable uses (Hanney and others in press). Furthermore, it would be important to check whether, for example, costs are not simply being transferred elsewhere. Improvements could arise also in the process of health care delivery; these could be measured by techniques such as patient satisfaction surveys.

Broader economic benefits

A range of benefits can accrue to the national economy from the commercial exploitation of research. These can take the form of employment and profits resulting from the manufacture and sale of drugs and devices (Rosenberg 2002). The national economy could benefit also from exports and/or import substitution (Gadelha 2000; Hale and Towse 1995).

While there is a danger of double counting, it may be valuable to adopt a human capital approach and focus on the value of production that is gained from having a healthy workforce. This can be measured by examining the reduction in days off work. Typically, a human capital approach has been used, in which potential future earnings are calculated for people who, as a result of advances in medical research, can continue to contribute to national production; for example by reducing the days off work caused by low back pain (Drummond and others 1992; Mushkin 1979; Weisbrod 1983). However, those who use it share concerns that such an approach could have equity implications in assessing the benefits, in that it would seem to favour research that is relevant for those of working

age. This concern might be relevant here in that many who suffer most from arthritis are retired. An additional concern is the value of the production lost from days off work may be overestimated if it is measured in terms of relevant wage-rates (Koopmanschap and others 1995).

2.1.2 Payback model

The second element of the evaluation framework is the logic model and its various stages are shown in Figure 2.2. The interviews and subsequent review of the model by the Development Committee confirmed its broad structure although, as with the payback categories, some of the stages were relabelled in order to be applicable to the wide portfolio of research that is funded by arc.

The linearity of the model serves to indicate a series of assessment stages and does not claim to represent exactly how the research translation process necessarily, or usually, works. Particularly in relation to more basic research, the initial flows of outputs from the research are into the pool or stock of knowledge and from there back to further research. While it is not completely possible to tie the categories of benefits to certain stages of the model, it is possible to identify broad correlations: the *knowledge production* and *research targeting and capacity building* categories together are generally the primary outputs from research; the *informing policy and product development* category relates to the secondary outputs; and the categories *health and health sector benefits* and *broader economic benefits* respectively are generally the final outcomes. Hence, although each category of output was assessed for each stage of the model, certain stages tended to produce certain outputs.

Stage 0: topic/issue identification

This stage involves the generation of the original ideas for the research and varies considerably, depending on whether the main driving force is generated internally by the researcher or generated externally. Most arc funding falls into the former category; for many researchers the topics will be curiosity-driven and based on the researchers' examination of the existing stock or pool of knowledge and opinions about where gaps and/or opportunities exist, and where further research could advance understanding. Such factors will inform more clinical and AHP researchers but here, consideration of clinical needs could be a factor also and might be based on personal experience of treating patients. Where research topics are generated externally, the identification of the issue comes from a process of needs assessment that could involve analysis either within the scientific community or more widely. In the latter case, many groups could be involved, including not only members of the wider research community and representatives of research funding bodies, but also potential users and beneficiaries of the research that are drawn from some combination of the wider political, professional, industrial and societal environment.

Interface A: peer review

The nature of the activities at Interface A will vary depending on the type of issue identification. Where the issues are generated internally, the interface involves traditional processes whereby the researcher develops a detailed proposal and submits it for peer review. In fact, there is not really much of an interaction between researchers and the wider environment in this case as often, peers are required to assess on the quality of the science (although the issues of relevance and impact may be influential also). Where the topics are generated externally, the interface issues become more important as there are potential difficulties in ensuring that the research community is actively engaged with the priorities that have been identified, and that the project specification meets the needs as

identified. In both cases, however, there are issues about how far the proposal was subject to changes as a result of the review process.

Stage 1: inputs to research

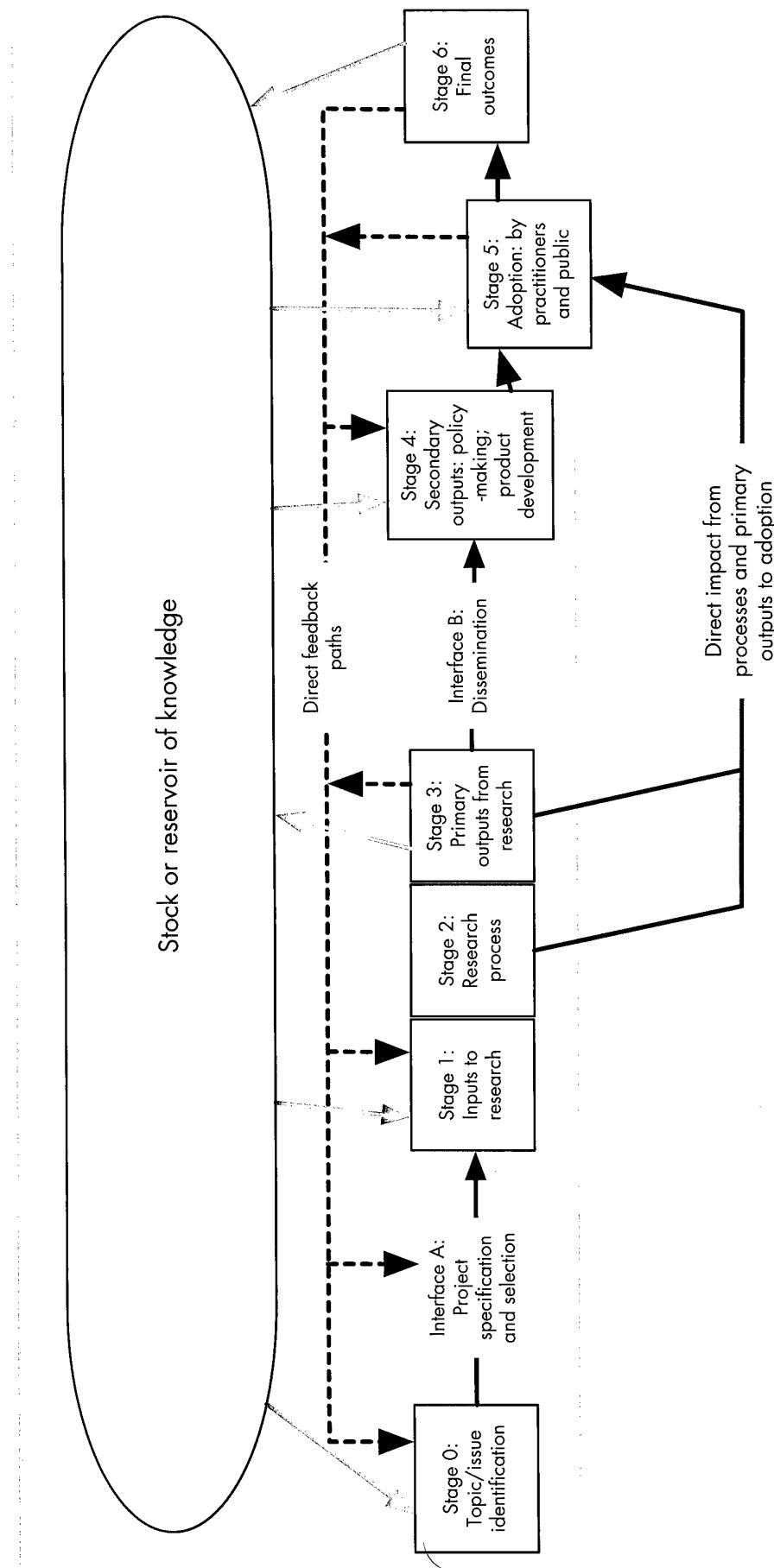
It can be important to consider not only the financial inputs, including any beyond the specific arc funding, but also the expertise of the research team and the knowledge base on which they have built. As with looking at the sources of follow-on funding, the idea behind examining other funding brought in to support arthritis research is to look at whether arc funding is helping to increase the funding of arthritis research by general funders of health research. One way it could do this is by arc giving its support to studies which produce findings that other funds believe are worthy of further investigation.

Stage 2: process

Consideration can be given as to how appropriate the proposed methods turned out to be, and whether any difficulties were encountered. In some cases it could be relevant to explore how far potential users were involved at this stage. It is possible that the difficulties identified at this stage could explain later problems with translation or uptake of the research findings.

Stage 3: primary outputs

Knowledge production as represented by the various types of publications is a major primary output from the research (various ways of measuring this were discussed above). Most of the primary outputs will feed into the stock of knowledge and become part of the body of knowledge that informs further research or is incorporated into systematic reviews. The research benefits in terms of targeting future research and capacity building can be seen as featuring here also, but they represent either feedback to further research that is conducted by team members, or findings that feed into the stock of knowledge and help target future research. Further, we may consider how far the career development that is based on arc funding helps to propel some researchers into positions within the health care system, where they can play a role in ensuring that the later stages of translating research findings into outcomes are achieved.



The political, professional and industrial environment and wider society

Figure 2.2: Payback model adapted for the arc study

Interface B: dissemination

Usually, dissemination is seen as being somewhat more active than the mere production of academic publications. However, there are clear overlaps between some activities, and sometimes it is possible to record not only dissemination activities but also the successful transfer of research findings to potential users in the political, industrial and professional environment and wider society. Presentations to potential academic and user groups and media activities are major ways of disseminating findings, as is the production of brief summaries of findings which are targeted at specific user groups. In previous case studies attention has focused also on the way in which some researchers conduct study days or training that is based on the approach developed by their research: these can be highly effective dissemination mechanisms.

Stage 4: secondary outputs

A wide range of items can be considered to be secondary outputs. In terms of policies, the key issue is that policymaking involves those in positions of authority making choices that have a special status within the group to which they apply. The results can take many forms, ranging from national health policies made by the Government to clinical guidelines determined by professional groups, to guidelines or care pathways, etc. that are agreed within local units. Clinical guidelines provide a particularly fruitful form of secondary output on which to focus analysis (Grant and others 2000). As noted above, many other items can be included also if they are informed by research findings, for example, "how-to" manuals, criteria adopted by evaluative or inspectorial bodies, training packages and official curricula, legal decisions and media campaigns by health care providers. Where the research seems to have resulted in secondary outputs, it is useful to explore the factors that have led to this.

The position of systematic reviews is more complex. They are themselves a form of research and, where commissioned, can be considered using this framework, but inclusion of a study in a systematic review is a form of secondary output and might lead on to further use.

In relation to product development, it is important to distinguish between the commercialisation of a proven approach (which is a secondary output) and consideration by a drug company of therapeutic potential or novel mechanisms (which is a primary output). For example, if research findings are directly built upon in the process of developing a commercial product, such as a new drug that is licensed for arthritis, this can be seen as an important secondary output. For example, arc-funded research played a key role in the introduction of anti-Tumour Necrosis Factor (anti-TNF) therapy for arthritis.

Stage 5: adoption

Behavioural change by practitioners and/or the public is necessary for the research findings incorporated into secondary outputs to result in final outcomes. This may involve take-up of new drugs or procedures as set out in a secondary output such as a guideline from the National Institute for Clinical Excellence (NICE). Sometimes the adoption comes as a direct result of the primary outputs, as when clinicians – often at the cutting edge – decide to implement research findings even prior to the development of clinical guidelines. Eitherway, it is important to try to establish the adoption or take-up rates and to explore how far the behavioural change can be attributed to the specific research findings, as opposed to other factors such as a more general change in climate of opinion in relation to, for example, the importance of exercise.

The role of the public in responding to informed advice – which is often research-based – is seen as increasingly important, especially in a field such as arthritis (Leong and Euller-Ziegler 2004). Various factors can be explored here, including the extent to which patient behaviour might change as a result of interactions with health care providers who promote research-based messages, and how far the public might respond directly to publicity about research findings when they are used: for example, in media campaigns encouraging participation in preventative activities.

Stage 6: final outcomes

The final outcomes are the health and broader economic benefits identified in categories D and E in Box 2.1. Increasingly, these are seen as being the ultimate goal of health research funding, but their precise estimate in practice often remains difficult. At one level there might be data such as audit figures available from areas where there is known to have been local implementation of the research findings. At an overall level it is possible that figures would be available for the potential population who could benefit from the new drug or procedure and information about the level of benefit that a patient might receive. If knowledge about adoption levels then was taken into consideration also, it might be possible to indicate the levels of benefit.

2.2 Case study selection

A case study approach has been used, and recommended for future use, where the emphasis is on showing the long-term benefits from health research (Lavis and others 2003). More widely, there is a long history of applying the case study approach to examine the translation of research (Yin and Moore 1988).

Case studies enable narratives or stories to be told in order to illuminate how the research funded in the early 1990s was translated (or not) into practice; thus each case potentially provides an illustrative example of the long-term outcomes from **arc** research. Within a case study approach, selection of cases generally does not follow a straightforward sampling logic in which those selected are assumed to be representative of a larger group (Yin 2003). A study adopting a multi-case approach such as this is aiming to ensure that the benefits from the full range of modes of funding and types of research can be illustrated, and that there is scope for some cross-case analysis. The selection of cases is, therefore, somewhat purposive. In their classic case-study analysis of research utilisation, Yin and Moore (1988) went to considerable lengths to ensure that they were including only studies where utilisation was thought to have been. We did not go that far, but given that one aim of the research was to illuminate the long-term outcomes, we considered it sensible to concentrate on studies where it was judged that there was a reasonable chance of a range of research paybacks.

In deciding the time window to use for selecting case studies there were two competing factors. On the one hand, we needed a sufficiently historic time window to allow the opportunity for the outputs of the research to be translated. On the other hand, we needed to collect evidence – both from archival material and interviews – which would not be biased by difficulties recalling work that had been undertaken some time ago. **arc** instituted a new computerised database during the early 1990s and all the grants that have been awarded since 1990 are held on this database. Prior to this, only paper records of unknown completeness were available. Therefore, we decided to select grants that were awarded between 1990 and 1994. In the event, the records available and PIs' recall of events was very good for this timeframe; however, as discussed elsewhere, some of the

outputs of the research have not yet come to fruition; this was particularly the case with the more basic research.

With only 16 case studies we could not hope to for a group that was representative of **arc** funding in any statistical sense. Our case study selection procedure was a trade-off between the risk of biases and the risk of ending up with case studies where few outcomes had occurred or where we could not investigate the case study, for example, if the case study PI was unwilling or unable to cooperate with our study. Inevitably this leads to concerns about the possibility of generalising of our conclusions. With each of our conclusions and recommendations we have taken care to consider how far our findings can reasonably be generalised, and to word them in the light of this consideration.

There were three stages to case study selection:

1. generating a selection matrix to ensure that we selected case studies which reflected the variety of **arc** funding;
2. populating this matrix with suitable grants; and
3. selecting individual grants from this matrix.

2.2.1 Generating a selection matrix

We identified three key characteristics of each grant: the type of funding (see Box 2.2), the area of research and the PI's level of success. Our original selection matrix, shown in Figure 2.3, had 16 cells and concentrated on the four main types of funding, two areas of research (clinical science and basic science) and two levels of success ("High" and "Mid"). A summary of how we classified grants is given in the following sections; a detailed description, including the bibliometric techniques, is given in Appendix A.

We chose to classify according to funding mode as this is an obvious policy lever for **arc** in affecting the outcomes of the research that it funds. We chose to distinguish between two areas of research: basic and clinical, as we felt that they might be expected to produce different spectrums of feedback and, again, could be influenced by **arc**. Further, we wanted to examine researchers of differing levels of success to see whether the amount or pathways of translation for the types of researcher differed. The final reason for classification was to ensure that the case studies selected reflected the variety of grants that **arc** awards.

Mode of funding	
<i>Project grants:</i>	These grants provide support for a project designed to seek an answer to a single question or small group of related questions. Project grants are of limited duration, usually up to a maximum of three years, and may provide for the salary of postdoctoral or technical assistance, running costs and the purchase of small items of essential equipment.
<i>Programme grants:</i>	Programme grants are awarded to established groups undertaking research relevant to arc's aims. They provide longer-term support, where the aim is to answer an interrelated set of questions. Programmes are potentially renewable and may provide funding over five years for the salaries of academic and technical staff, plus running costs and small items of essential equipment
<i>Fellowships:</i>	Fellowships are awarded to attract and retain talented clinical and non-clinical scientists in research that is relevant to arthritis. For the current study we looked at clinical research fellowships, junior research fellowships, non-clinical research fellowships and senior research fellowships. These fellowships are for a mixed duration but provide salary for the PI and associated research costs.
<i>Institutes:</i>	arc provides core support for its two research institutes, the Kennedy Institute of Rheumatology in London and the Epidemiology Research Unit (arc ERU) in Manchester. This support underpins all their major scientific and clinical activities.
<i>Type of researcher:</i>	As described in Appendix A, we used qualifications listed in the grants' database to classify PIs as basic scientists, clinical scientists or AHPs.
<i>Bibliometric impact:</i>	As described in Appendix A, we identified a group of "high-impact" researchers and a group of "mid-impact" researchers based on the analysis of publication output and impact of career.

Box 2.2: Explanation of major grant characteristics

	Basic science	Clinical science
Project grants	High success	High success
	Mid-success	Mid-success
Programme grants	High success	High success
	Mid-success	Mid-success
Fellowships	High success	High success
	Mid-success	Mid-success
Institute funding	High success	High success
	Mid-success	Mid-success

Figure 2.3: The original selection matrix

During our key informant interviews it became clear that there were problems with our original selection matrix:

the area of research carried out by AHPs, which was taking up an increasing proportion of arc research funding, was not similar enough to clinical science research to be considered within that category;

in the case of work by AHPs, there were some funding types where no grants had been awarded;

the two institutes that arc funded carried out research of very different types, with the Kennedy Institute concentrating on biochemical and cell biological research and the arc ERU specialising in epidemiology.

These developments meant that our selection matrix evolved into the one shown in Figure 2.4. Unfortunately, we were not able to carry out more case studies than the 16 that had been originally envisaged. This meant that we worked with the Development Committee to select which cells of the matrix were most important; these are shaded in light blue. The selection was guided by three criteria:

1. trying to have at least four case studies in every group to compare – for example: four project grants to compare against four programme grants; or four “Mid” impact grants to compare against four “High” impact grants. This is analogous to experimental replication;
2. balancing the number of case studies against the level of funding committed to different areas and the number of grants awarded in those areas – for example: in our time window, AHP grants was the smallest area of funding and many of the fellowships awarded provided less than £1,000; consequently we only carried out two case studies on AHP work, both of which looked at project grants;
3. trying to avoid case studies that looked at the same subject area where possible – in one case (“Mid” Impact Clinical Project and “Mid” Impact Basic Manchester) we failed to achieve this and there are two case studies looking at very similar areas of research.

		Basic science	Clinical science	Allied health professional science
Project grants		High	High	High
		Mid	Mid	Mid
Programme grants		High	High	
		Mid	Mid	
Fellowship grants		High	High	High
		Mid	Mid	Mid
Institute funding	Kennedy Institute	High	High	
		Mid	Mid	
	arc ERU	High	High	
		Mid	Mid	

Figure 2.4: The final selection matrix

2.2.2 Populating the selection matrix

To generate shortlists for each cell of the selection matrix, we performed the following:

exported a list of all grants awarded between 1990 and 1994 from the arc funding database;

selected and classified grants by funding type;

classified the grants by area of work, using the PI's qualifications; and

examined each PI's publication record to classify grants by success.

The process of populating the matrix was more complex than we had expected, mainly due to the level of sophistication required in the assessment of success by bibliometric measures.

Complete list of arc grants from 1990–1994

arc had almost complete computerised records for all grants and applications since 1990. This database contained information on the size of grant that had been awarded and paid out, names of co-applicants, the title of the grant, the PI's address, the type of funding and the location of archive records of the grant. This database made generating a list of all grants awarded between 1990 and 1994 very easy. This generated a list of 632 grants.

Selection and classification by funding type

The arc grants database contained information on the funding type of each grant, and we used this information to classify candidate grants by funding type. In the early 1990s there were 14 types of grant;⁴ we chose to concentrate on the four most significant funding streams and to group all fellowships into one category (see Box 2.2 for explanations of our categorisations). We chose to concentrate on the types of grants that make up most of arc's current funding and to look at the types of funding which are most likely to be affected by policy changes. This generated a list of 556 grants held by 357 PIs.

⁴ The types of grants awarded between 1990 and 1994 were: academic post, building grant, clinical research fellowship, core support grant, educational project grant, equipment grant, junior research fellowship, non-clinical career development fellowship, PhD studentship, postgraduate bursary, programme grant, senior research fellowship, special purpose grant and travelling fellowships.

Classification by area of work

We classified the grants into the three areas of work: basic science, clinical science and AHPs, on the basis of the PI's qualifications between 1990 and 2001. This classification was reviewed then by the *arc* Scientific Secretary. With 357 PIs to classify we needed a technique that could be applied rapidly and that was reasonably clear cut. We did not think it was practical to classify grants from their titles alone and unfortunately, the *arc* grants database did not contain the grant abstracts that would have allowed a more nuanced classification.

Classification by success

This was the most complex part of shortlist generation. We assessed success in two steps: first, by the examining publication record of each PI; and second, by ranking the PIs working in each area according to their publication record. We scored publication record in three different ways and generally used the top decile of researchers for the "High" and the mid-tenth (from the 45th to 55th percentile) for the "Mid" group. We also tried to select researchers who were ranked similarly by all three methods of scoring. In some cases we had to widen these definitions, notably for AHPs and programme grants, where the number of grants was very small. We chose to focus on the "High" group as we hoped that their grants would be the most likely to show evidence of translation. The "Mid" group was picked as they are similar to many of the researchers that *arc* funds, due to the non-linear distribution of publication record. In addition, we were keen to avoid the political difficulties of recruiting researchers that we had identified as "Low" success.

We examined the publication output of each PI from 1990 to 2001. We assessed PIs rather than grant output as we did not have the necessary computerised records of the output of the 556 grants. Our only record of publications attributable to grants was the end of year reports supplied to *arc* by the PIs (published as *arc* Scientific Reports). These records only covered the period of the grant; so did not include publications submitted after the grants had finished; furthermore, from previous work we had reason to suspect that these records would be incomplete and inaccurate. Previous attempts to examine the outcomes from specific pieces of research have encountered difficulties in obtaining precise publication records from researchers, partly because many researchers see individual studies as part of a wider stream of work (Buxton, Hanney and others 1999; Hanney, Soper and others 2003).

City University's Bibliometric Research Group was commissioned to identify each PI's papers by searching the ISI Science Citation Index (SCI) for all papers matching their name and initials. We then filtered out the relevant PI's papers in a novel multi-step approach that is explained in detail in Appendix A. The "success" of each publication was measured by the Journal Impact Factor (JIF) of the journal in which it was published. The JIF is the average number of citations received by publications in the journal. Because we had identified over 15,000, it was not practical to check the actual number of citations received by each publication.

2.2.3 Selecting the individual grants

The final selection of case study grants from the shortlist was carried out by the Development Committee using their knowledge of the field, the individuals involved and the variety of *arc* funding. In the early stages of the project we asked the members of the Development Committee to score, on a scale of 1 to 5, the shortlisted case studies before the selection meetings; scoring would be done on the basis of how interesting and useful they thought each grant would be as a case study. These scores allowed us to produce a reduced shortlist of case study grants. This list contained the four or five grants that had

been scored the highest, for each cell of the selection matrix. The Development Committee then selected the case study grants from this list. In the later stages of the project the Development Committee selected directly from the complete shortlists.

2.2.4 Approaching PIs

After the Development Committee had selected candidates for case studies we approached the PIs by email and telephone. All the PIs that were approached consented to their grants being examined for this study.

2.3 Data collection/methods

For case studies it is appropriate to use multiple sources of evidence relevant to the same issue and to adopt a process of triangulation, whereby a number of partial sources that point towards the same conclusion are used to increase our confidence in that conclusion. To construct the narrative of **arc**-supported research for each of the 16 case studies, three independent sources were used: document and literature review; semi-structured interviews; and bibliometric analysis, these are described below. As this included the possibility of interviewing NHS employees, we had to seek ethical approval to carry out the study and this was obtained from the Multi-Region Ethics Committee (MREC) for London (reference MREC/03/2/059).

2.3.1 Document and literature review

We reviewed a range of material available from **arc** including, for example, original grant applications; referees' reports; subsequent correspondence between **arc** and the grant holder; and the review of interim and end of grant reports. In addition, case study analysts read the peer-reviewed papers and reports that were considered by the PI to have arisen from the **arc** grant and other background literature.

2.3.2 Semi-structured key informant interviews

We interviewed PIs, named and unnamed staff on awards, collaborators and other stakeholders, including peers and users of the research. The interviews were based around the payback framework and explored the origins of the research, primary outputs and any translation of research findings into product development, policy and practice. Most interviews were conducted face-to-face, with a few by telephone. In total we interviewed 42 individuals,⁵ with each interview taking between 60 and 150 minutes. In addition, interviews were often followed up with email correspondence and telephone enquiries. All the interviews were recorded and these recordings were preserved for later reference.

2.3.3 Bibliometrics indicators

The bibliometric indicators were commissioned from City University's Bibliometrics Research Group.⁶ The source data is the CD-ROM version of the SCI for 1990–2003 and the Research Outputs Database (ROD) for 1990–2002 (see Dawson and others 1988 for more detail). ROD itself is a subset of all UK biomedical articles, notes and reviews downloaded from the SCI and Social Science Citation Index, with information on postcodes and acknowledged funded sources added. Below, the indicators used in each of the case studies are defined. They are split into two sets. The first, "Publication Portfolio" indicators, describe the papers arising from the funded **arc** research study. The second,

⁵ This excludes the five key informants we interviewed in the development stage of the project.

⁶ See <http://www.city.ac.uk/researchprospectus/informatics/information.htm> for more information.

“Knowledge flow” indicators, described how others use those papers – that is, how often the papers are cited and what the characteristics are of the citing papers. To identify papers arising from a particular research grant, we compiled a long list of papers; these were identified by searching the SCI and other bibliographic sources for the PI’s name over the period of the grant and the five years after the grant had ended. In interview the PI was asked to review the list of papers and make amendments as required. The revised shortlist of papers was deemed to be “attributable” to the grant.

Publication portfolio indicators

Output (number of papers) – the number of peer-reviewed papers published in the serial literature as identified by the PI and “attributed” to the case study arc research grant.

Collaboration (mean number of authors, mean number of addresses, proportion of papers with a non-UK address) – the average number of authors and addresses on papers attributed to the arc research grant, and the percentage of papers with a non-UK address. This indicates the degree of individual (author), institution (address) and international (non-UK address) collaboration.

Type (research level distribution, proportion outside “ARTHR” field). The distribution of papers attributed to the arc research grant by research level. The research level is a journal classification system developed by CHI Inc., which is based on expert opinion and journal-to-journal citations and which has become a standard tool in bibliometric analyses (Narin and others 1976). Journals are allocated to four hierarchical levels, in which each level is more likely to cite papers in journals at the same level or the level below it, and vice versa. Typically, only 4% of the papers in level 1 “clinical observation” journals (eg *British Medical Journal*) will cite papers in level 4 “basic” journals (eg *Nature*), compared to 8% for level 2 “clinical mix” journals (eg *Arthritis and Rheumatism*), and 21% for level 3 “clinical investigation” journals (eg *Immunology*).⁷

The “ARTHR” field is a bibliometric definition of arthritis research, based on key word searches and specialist journals, that aims to identify all arthritis research papers (Lewison and Devey 1999). Attributable papers from arc outside the ARTHR field indicate research that will be of general interest to a number of research areas and that is more likely to be fundamental or basic in nature.

Funding (mean number of funders; proportion acknowledging arc) – the average number of funding acknowledgments on papers attributed to the arc research grant, and the proportion of those papers that explicitly acknowledge arc. The funding body acknowledgements are derived from ROD. This indicates the degree of additional funding of which the investigator is in receipt.

Knowledge flows

Strength (mean number of citations per year from beginning of grant to 2003; mean number of citations for the five years following publication of the paper) – the average number of citations to papers attributed to the arc research grant for

⁷ In addition to presenting the CHI Research Levels, in Annex A of Volume II we provide research levels calculated on the basis of keyword searches (RL-calc). This is a new method developed by Lewison and Paraje 2004 which has the advantage of providing a paper-specific description of the type of research (as opposed to a journal specific description resulting from the CHI system).

different timeframes. Both indicators provide a proxy measure of how the research has supported other published research.

Knowledge translation (research level of citing papers,⁸ relative research level) – the distribution of research levels for papers citing those papers attributed to the arc research grant by research level. This indicator assesses whether, for example, basic research is being cited by clinical research. The relative research level is the research level of the cited paper minus the research level of a citing paper. A positive value is an indicator of forward translation (that is from basic research into clinical); a negative value is an indicator of backward translation (from clinical research into basic).

Knowledge diffusion (proportion of US cites; proportion outside “ARTHR” field) – the proportion of citations coming from papers with a US address. The US is selected as it is the largest scientific producer (accounting for approximately 40% of biomedical research papers) and provides an indicator of the diffusion of arc-funded research outside the UK. The proportion of citations arising from papers which do not fall within the bibliometric definition of arthritis research (ie ARTHR bibliometric search) provides an indicator of the diffusion of arc-funded research outside the field.

Each of these indicators are presented in a standard form for each of the case studies published in Volume 2 (Wooding and others 2004). In addition, all the bibliometric indicators are listed in Annex A of that volume. In presenting our results in Chapter 3, we put these indicators into context by making comparisons with all the 16 case studies, all the papers resulting from researchers who were supported by arc between 1991 and 1994 and, where the information is available from secondary sources (such as Dawson and others 1998; Lewison and Devey 1999; Wellcome Trust 2001), arthritis research and biomedical research in the UK.

2.3.4 Clearance and validation

In every case a draft copy of the case study report was sent to the PI as a professional courtesy and for comment and approval. All 16 case studies were cleared for accuracy by the PIs, with 15 giving approval for publication.

2.4 Case study pilots

Given the novel aspects of parts of this research, we undertook three case study pilots and reported these to the Development Committee before proceeding with the subsequent 13 case studies. The main issues that emerged from the pilot studies were the importance of interviewing the PI early in the process, prior to extensive review of subject specific documentation, and of ensuring a common use of language within the payback framework.

2.5 Cross-case analysis

Using the information collected from the document and literature reviews, semi-structured key informant interviews and bibliometric analysis, each of the 16 cases was

⁸ Calculated Research Levels (RL-calc) are provided in Annex A of Volume II.

written up as a narrative, organised according to the structure provided by the payback framework. Using a common structure has the advantage of facilitating comparative analysis, allowing us, for example, to identify the factors associated with the successful translation of research. We employed two approaches to our cross-case analysis. The first was based on a qualitative assessment of the case studies based on a discussion of the key observations made by each of the project team members. The second involved a novel method of scoring the case studies along the five arc payback categories. Both are described below.

2.5.1 Qualitative analysis

Each member of the project team was asked to review all 16 case studies and, prior to the workshop, submit their observations relating to the policy, evaluation and research implications arising from the case studies (see Box 2.3). These observations were then collated and analysed to allow the main themes emerging from the case studies to be identified. Each theme was taken in turn and, in group discussion, the main issues and sub-issues were discussed and validated with supporting and/or contradicting case study evidence.⁹ It is these themes – supported by the quantitative assessment described below – that form the basis for the key policy observations and recommendations that are reported in Chapter 4.

Observations about the process of arc-funded research and the development of that research, eg: "Where translation occurs, it occurs largely due to the personal networks of the PI."

Observations about how arc monitors and evaluates its research, eg: "The arc does not have a comprehensive method of monitoring publication outputs that occur after the end of the grant."

Observations about evaluating research in general, eg: "PIs with many grants tend to run their laboratories as 'one pot' of money, hence attributing outputs is difficult."

Box 2.3: Observations requested from workshop participants

2.5.2 Quantitative analysis

To allow comparisons between groups of case studies we developed a method for converting the qualitative descriptions of the payback in each of the five categories into a more quantitative measure. We did this using a novel system of scoring that drew upon the scoring system used in the RAND/UCLA Appropriateness Method (University of California at Los Angeles; Fitch and others 2001). We felt that the key issue was consistency: would different individuals score the case studies similarly when working from the qualitative descriptions of outputs?

⁹ See McFadzean (1998) for further details.

Scoring method

We carried out the scoring process, working with all nine members of the project team during the course of a two-day workshop. Prior to the workshop we prepared a summary table of the payback from each of the case studies in each of the categories (presented as Table 3.1 in Chapter 3). Working through the payback categories, we asked all nine team members to individually score the level of payback from each case study in each category. Scoring was done on a scale of 1–9. At the bottom of the scale, 1 represented no payback. At the top end of the scale the value of 9 was a hybrid of two definitions. It was informed by both the size of the largest payback seen in that category within the 16 case studies but also reflected the maximum payback that could have been expected from the grant. This meant that a well-executed, successful clinical trial could score highly for knowledge generation, even if the area of knowledge to which it contributed was relatively small, as it had achieved all that could be reasonably be expected.

The scores from the first round were analysed and sheets showing their distribution were circulated to the team. Different sheets were provided to each team member; a portion of such a sheet is shown in Figure 2.5: the case studies are shown by the letters A–E (not all of the case studies are shown), the distribution of scores is shown by the histograms and the team member's score is shown by the shaded bar. Following this, there was discussion about scoring discrepancies in order to resolve any misunderstandings between the team about the outputs from the case. The team members then had the opportunity to rescore all of the case studies for that payback category, so providing a set of second-round scores. The median of these second-round scores has been used for the analysis presented in this report. By revealing only the distribution of the first-round scoring (rather than individual scoring) the likelihood of outliers feeling pressured into scoring closer to the group consensus is reduced.

Conversion of qualitative data into quantitative data is a valuable tool to allow comparisons; however, in any system for doing this there will be an inevitable loss of information and opportunities for subjectivity. The main areas of concern with the novel system that we used regarded:

- the differences in quality of the case study evidence as a consequence of having been carried out by different researchers;
- the different levels of understanding of the case studies and their paybacks due to the varying expertise of team members and their differing familiarity with the case studies;
- the subjective allocation of a score based on the qualitative information provided;
- the possibility of systematic bias, for example, that the scoring approach could favour focused research that produced specific outputs and outcomes.

We hoped to get a measure of the level of subjectivity in the scoring by using a group of individuals and then by assessing the level of agreement within that group. We also examined whether the researcher who had carried out a case study tended to rank it differently to other members of the project team.

Determining levels of disagreement and agreement

The RAND/UCLA method provides an algorithm for determining whether a set of scores disagrees; however, this method is relatively insensitive to disagreement as the RAND/UCLA method is most concerned with picking up the occasions where people are scoring at opposite ends of the scale. In addition to the RAND/UCLA disagreement

algorithm we also tested agreement by checking the number of scores more than one and two units away from the median score (these results are discussed in Chapter 3).

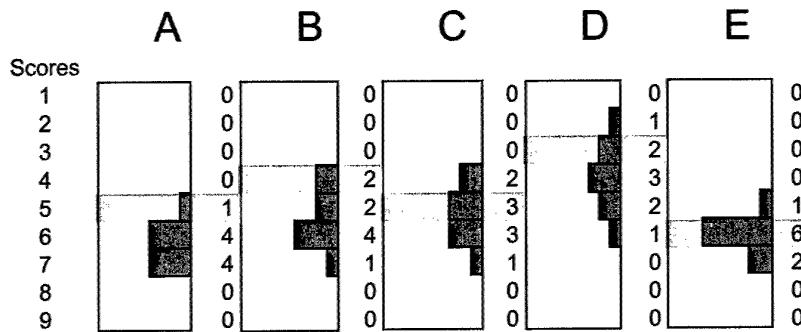


Figure 2.5: Feedback provided to team members during case study scoring

Displaying scores

To provide a visual representation of the levels of payback in each category, we developed “payback profiles”. We plotted the median payback scores for each category of each case study on one axis of a spider or radar plot to produce pentagons whose size and shape represented the payback from the case study. By using a transparent fill for the pentagons and overlaying the profiles for a group of case studies, we can glean an impression of the level of similarity between case studies and where payback is occurring (as illustrated in Figure 2.6) showing how the profiles of groups of case studies can be compared. Four payback profiles are shown on each set of axes.

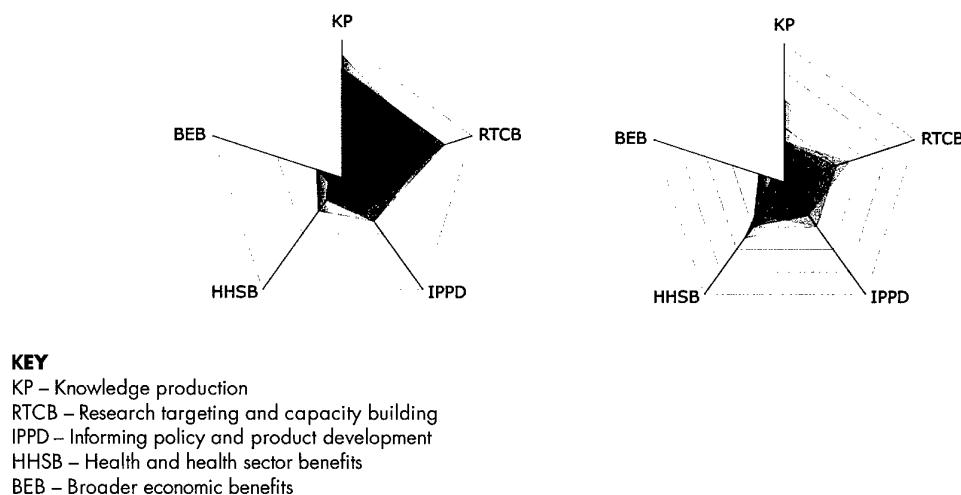


Figure 2.6: Two groups of “payback profiles” generated with artificial data

The preceding chapter explained how we proposed to evaluate the long-term outcomes of research supported by the Arthritis Research Campaign (arc). It paid particular attention to the development of an evaluation framework that included the development of the arc Payback Categories and Model. In addition, we explained how we selected 16 case studies from a sample of 556 potential grants, and the data collection methods that were employed.

In this chapter we present the results of our evaluation. We begin by cataloguing the diverse range of research outputs and outcomes that were identified. To help our analysis we sort these by the five payback categories – knowledge production, research targeting and capacity, informing policy and production development, health and health sector benefits and wider economic benefits – and present summaries both in a qualitative and quantitative form. Finally, we compare differences in the various groups of case studies; for example, we compare payback for project grants versus programme grants.

In the following chapter we draw out a number of policy observations that arise from the data and analysis presented here, and then make a number of recommendations for arc.

3.1 An analysis of payback categories

In this section we catalogue the research paybacks identified from the 16 case studies, as summarised in Table 3.1. Below we assess these results by the five payback categories. As noted in the methods section, these profiles are a novel way of synthesising these data and they should not be over-interpreted.

3.1.1 Knowledge production

In total we identified 302 papers attributable to the case study grants, and between the beginning of the case studies and 2002 these received a total of 975 citations per year. The major themes covered in these publications are summarised in Table 3.1. The researchers funded by the case study grants had presented their results also at numerous national and international conferences; however, as there were no consistent and comprehensive records of these presentations we have not collated information on them. The largest number of papers were produced by case studies E and L, over 40 papers each. The highest levels of citation were achieved by case studies C, H and L, with the papers from these case studies all receiving over 100 citations per year; for case study L, on anti-TNF, over 300 citations per year. Continuing its presence at the top of the bibliometric indicators, case study L joined case study C with eight or more citations per paper per year, the highest level among the remaining case studies being five citations per paper per year.

In considering measures of knowledge production based on citation it is important to remember the disadvantages suffered by the allied health professional (AHP) work. Of the papers attributed to these grants, case studies A and I, only 53% were published in journals for which citation information existed in the Science Citation Index (SCI); furthermore, many of the citations of the articles that are included will be missed, as they will be from journals not covered by the SCI.

In comparison to wider selections of researchers (see Table 3.2) it is notable that the average level of citation is markedly higher – as might be expected, given our selection technique of picking “High” and “Mid” impact researchers. The other notable differences are in the distribution of papers by research level. Our case studies produced a similar number of basic science journal articles to the overall “ARTHR” field, but fewer than either the pool of grant holders from 1990–1994 or the overall output of our case study researchers. This lower representation of basic articles is mirrored by a concomitant increase in the number of articles in “clinical mix” (CHI research level 2) journals. The fraction of articles in non-classified journals is highest for the articles attributable to our case studies and this probably reflects the two allied health professional (AHP) case studies.

3.1.2 Research targeting and capacity building

The majority of case studies recorded significant paybacks in terms of research targeting and capacity building. A major element was the career development of either the principal investigators (PIs) or postgraduate and postdoctoral research assistants who were employed on the research grants. For example, 24 scientists were awarded either a PhD or MD as a result of the research undertaken on the 16 grants, and for case studies I and J these were awarded to the PI. A number of the PIs were appointed subsequently to senior academic positions including research chairs, clinical consultants, heads of Medical Research Council (MRC) units, the dean of a medical school, etc, but it is inappropriate to attribute wholly such outcomes directly to the arc funding. That said, in our interviews it was widely acknowledged that arc support at least had an influence on such career development.

A further element of research capacity was the transfer of technological know-how – in case study O, genetic mapping from a site acknowledged as leader in the field which hosted the PI for a six-month training fellowship before returning to her host institution. Similarly, in case study P, a number of interviewees commented that the establishment of the Norfolk Arthritis Register increased (arthritis) research capacity locally.

Table 3.1: Outputs and outcomes of arc-funded research – payback from 16 case studies

Case	Knowledge production	Research targeting, capacity building	Informing policy and product development	Health benefits	Broader economic benefits
A	6 papers, receiving a total of 5 citations per year. The RCT concluded that an exercise class is more clinically effective than traditional GP management, regardless of patient preference, and is more cost-effective. Some papers describe the main findings, others discuss key methodological issues.	Further studies by the PI applied lessons from this study. The PI was involved also in the MRC-sponsored >£2 million UK BEAM Trial which draws heavily on the study. An interviewee thought that the project's success helped encourage arc to support AHP research. Provided research training for team. 1 PhD.	The work was not cited in any national clinical guidelines or equivalents. But it was cited both in a number of reviews, receiving high-quality assessments, and in a task group report. At a local level it is influencing policy in various places as part of a general move towards more active management of back pain.	The study showed that exercise was more clinically effective than GP management, with some improvements on disability questionnaire scores and back pain scales with less call on health care resources after the intervention. It is being applied in various places but sometimes to small numbers.	The study showed considerably fewer days off work in the period following the intervention for those in the intervention group than in the control group. Therefore, it would contribute to a healthy workforce. This is assumed to be happening in adoption sites but has not been assessed.
B	6 papers, receiving a total of 8 citations per year. The research showed that hip OA is a condition that arises through an interaction of a predisposition to the disease and mechanical insults to the hip. Hip OA is therefore an occupational disease in men whose work entails frequent heavy lifting.	The methods developed for the UK study were applied to a Japanese study on occupational lifting. All three of the co-applicants had chairs, and one of the two research nurses continued to work in the unit from where the work was undertaken.	Three papers were cited in three systematic reviews, each receiving high-quality assessment. One paper was cited in a Dutch clinical guideline and by the Industrial Injuries Advisory Council (IIAC) Assessment on whether hip OA in farmers was a prescribed disease.		Adoption of IIAC recommendation will result in monetary benefit to farmers suffering from OA, at a cost to the taxpayer, improving economic equity.
C	13 papers, receiving a total of 118 citations per year. A detailed understanding of the mechanisms of Matrix Metallo Protease (MMP) inhibition by Tissue Inhibitors of Matrix Metallo Proteases (TIMPs), coupled with information on the domain structure of the TIMP protein.	PI's industrial collaborator is now a senior research manager with Celtech. Grant supported assistant's PhD: the foundation of his career in pharmaceutical industry. Reagents produced have been sent out to over 100 researchers. MMPs initially seen as promising target, leading to drug development.	At least one drug development programme in industry continues, but no registered products have yet been developed.	No evidence of health benefits.	No evidence of broader economic benefits.
D	11 papers, receiving a total of 17 citations per year. Increased understanding of how the immune system deals with chlamydial infection and the causes of reactive arthritis. Three of the papers were peripheral to the grant subject and do not directly concern chlamydial infection or reactive arthritis.	The PI's career has continued to advance, but outside the field of arthritis. PI developed a mouse model of chlamydial infection, a method for identifying novel vaccine targets and possible candidate peptide for a vaccine against chlamydia, but this work was not developed further.	No evidence of informing policy and product development.	No evidence of health benefits.	No evidence of broader economic benefits.
E	44 papers, receiving a total of 73 citations per year. Referee commented in the programme renewal that "the major criticism is that no binding insights had come out of the programme". Site visitors felt that the team should publish in more international journals. PI contributed to two books.	Two postdoctoral researchers established themselves as researchers. A large number of current clinical leaders and several basic scientists trained on the programme. PI became leader in health services research (HSR).	Citation in one European guideline. No evidence of impact on policy. Some commercial interest in biomechanical processes in OA currently being trialled (although not directly linked to PI's earlier effort to interest footwear manufacturers in his findings).	No current benefit beyond small number of patients treated with the team's new taping technique. Subsequent research by team into screening using genetic and biochemical markers could lead to more effective and targeted treatment.	No evidence of broader economic benefits.
F	23 papers, receiving a total of 32 citations per year. Showed that collagen X was a definite marker for endochondral ossification and identified specific mutations in collagen X that cause chondrodysplasia type Schmidt.	Two research assistants obtained their PhDs because of the programme. Two of the postdoctoral researchers obtained positions at the university, while the third was offered a job offer at a prestigious organisation in the US.	The identification of specific mutations in collagen X was used to develop a diagnostic test for chondrodysplasia type Schmidt. Currently, this test is run as a clinical service for the European skeletal dysplasia network.	Confirmation of chondrodysplasia type Schmidt improves the management of care, and specifically prevents surgical intervention that may be applicable to other forms of chondrodysplasia.	No evidence of broader economic benefits.
G	32 papers, receiving a total of 56 citations per year. Showed that Complement plays an important role in processing immune complexes (ICs) and this may be related to SLE. The importance of anti-C1q antibodies in patients with SLE was also demonstrated.	PI established a successful research career and is in the process of setting up a medical school. Junior research assistant completed his PhD and is now an MRC supported senior research fellow. The other research assistants continue to work in the field of rheumatology.	The development of a diagnostic test that enabled the measuring of antibodies to C1q in SLE patients. The test is still used today at the Royal Postgraduate Medical School.	Prevalence of SLE is very low. The test allows diagnostics of one form of SLE (C1q deficiency).	No evidence of broader economic benefits.
H	20 papers, receiving a total of 108 citations per year. The PI and his team showed that the rheumatoid process appears not to be grossly abnormal, however its consequences are; it is this interaction between T cells and stromal cells that causes the problem.	MRC Centre of Immune Regulation has a major focus on rheumatology as a result of the fellowship. Fellowship initiated several subsequent project and programme grants. PI is now head of department. Academic collaborators are now leading scientists in the field. 7 PhDs.	The observation that RGD peptides were triggering apoptosis, informed the development of KGD peptides – a class of drugs used to treat deep vein thrombosis (DVT), but not useful for arthritis. Established long-term collaboration with industrial partners.	No evidence of health benefits.	No evidence of broader economic benefits.

Table 3.1: Outputs and outcomes of arc-funded research – payback from 16 case studies (continued)

Case	Knowledge production	Research targeting, capacity building	Informing policy and product development	Health benefits	Broader economic benefits
I	11 papers, receiving a total of 32 citations per year. Although the research hypothesis was not addressed in the timeframe of the fellowship, the research did contribute to a number of methodological developments.	The PI completed an MD and later became a consultant in rheumatology and general medicine. The lessons from the fellowship were applied to a subsequent project grant, assessing the effect of exercise on people with OA subsequently informed British and European guidelines.	A grading system for radiographic assessment of OA was developed and used in graduate clinical education.	It is now common for clinicians to recommend the physical training of muscles for people with OA.	No evidence of broader economic benefits.
J	13 papers, receiving a total of 10 citations per year. These publications showed that proprioception plays an important role in ACLD deficiency, and that a rehabilitation programme to enhance proprioception was more effective than traditional muscle strengthening.	The PI was awarded a PhD on the basis of the work.	Informed management of ACLD	Small benefit due to improved management of ACLD	Unquantified return in reduction of days off work
K	7 papers, receiving a total of 29 citations per year. Showed treatment with aspirin and heparin gives higher rate of live births in women suffering recurrent miscarriage associated with antiphospholipids antibodies than aspirin alone.	The trial generated a number of further lines of research, including the effect of antiphospholipid antibodies on <i>in vitro</i> fertilisation. The research registrar was awarded a MD. The PI established an international profile.	The British Royal College of Obstetricians and Gynaecologists guideline, along with US, Dutch and Australian guidelines, all recommend a combination therapy of aspirin plus heparin and all cite the trial as the evidence for this recommendation.	The trial led to the conclusion that a combined treatment increased the chances of a live birth for women who had tested positive for antiphospholipid antibodies. A number of these women would not have completed a pregnancy successfully without this treatment.	The treatment is complex and potentially expensive. However, as the number of live births increases, fewer costs are involved because of a decrease in repetition of pregnancies.
L	41 papers, receiving a total of 330 citations per year. The paper describing the first clinical trial demonstrated the efficacy and safety of treatment. The results of a four-centre double blind trial of 73 patients published in 1994 confirmed these findings.	Work promoted research in the area and promoted the use of biologicals in therapy. In wider RA research, work led to a paradigm shift, revealing that in RA's complicated multistep process, one single molecule can have profound effects on pathogenesis. 3 PhDs & MDs.	The research led to the development of three drugs which have been licensed in the UK. Anti-TNF treatment has been included in the guidelines of the NHS National Institute for Clinical Excellence and adopted by international consensus groups in Europe and the US.	In approximately 70% of patients, anti-TNF treatment leads to significant improvements in health. Health benefits might be limited because estimated costs per quality adjusted life year gained amounted to €50,000 (approximately £33,000).	During the first year of treatment, direct costs were reduced by 40% while indirect costs did not change substantially. The industrial partner has made substantial profits from drug sales.
M	15 papers, receiving a total of 20 citations per year. The research identified a number of specific genetic regions and markers for RA. By the end of 1996 170 sibling pair families had been identified, confirming linkage to HLA.	No evidence of research targeting or capacity building.	No evidence of informing policy and product development.	No evidence of health benefits.	No evidence of broader economic benefits.
N	16 papers, receiving a total of 75 citations per year. In the institute's submission for a new core grant in 1996, one reviewer noted that the work led to "few publications in first class journals". The PI suggested that in terms of papers the impact had been low.	The PI argued that in terms of papers the impact had been low, but the work had had a strategic influence on the field. The PI left the institute in 1994, which heralded the disintegration of the research team. He has subsequently refocused his research interests. 3 PhDs.	No evidence of informing policy and product development.	No evidence of health benefits.	No evidence of broader economic benefits.
O	17 papers, receiving a total of 21 citations per year. The core publications describe the identification of dinucleotide repeat polymorphism as the alpha-1-anti-trypsin locus and give a summary of the first 100 multi-case RA families with affected sibling pairs.	The fellowship benefited both the PI's career but also RA genetics capacity at the institute. 1 PhD.	No evidence of informing policy and product development.	No evidence of health benefits.	No evidence of broader economic benefits.
P	27 papers, receiving a total of 40 citations per year. NOAR has provided groundbreaking work in understanding how common RA is and the factors that are associated with it.	The PI, co-applicants and lead rheumatologist have become leaders in their field. Six people have obtained PhD/MDs as a result of their work on NOAR. NOAR has increased research capacity in Norfolk. The Danish have set up a similar registry, based on NOAR.	The incidence and prevalence figures from NOAR are widely cited in clinical guidelines. Recent data from NOAR on the benefit of early referral. Early referral of RA patients is one of the two quality standards set for rheumatology by the Royal College of Physicians.	No evidence of health benefits.	No evidence of broader economic benefits.

Table 3.2: Comparison of bibliometric summary data

	Papers attributed to case studies	Case study researchers' output from 1990–2001	Grant holders ARTHR field funded by arc in England between 1990 (1990–1997) ^a and 1994, examined from 1990–2001	ARTH field (1988–1995) ^b	Biomedical Science
Number of papers	302	1429	16503	6657	214364
Mean number of authors	5.6	4.3	4.2	4.4	3.8
Fraction with international addresses	38%	nk	nk	nk	nk
Mean JIF	23.7 ^c	13.7	12.7	nk	nk
Percentage of papers for each CII research level	1 2 3 4 nc	15 38 20 10 17	17 39 24 18 2	13 30 32 23 2	20 40 24 10 6
					15 22 22 29 12

a See Table A2 in Wellcome Trust (2001)

b See Dawson and others (1998)

c Figure given is mean number of actual cites in five years post publication, not JIF

In terms of research targeting case studies A, B, C, H, K and L seem to have the major impact. As summarised in Table 3.1. this ranged from informing:

- a >£2 million MRC-sponsored clinical trial (case study A);
- pharmaceutical companies' interest in recombinant TIMPs (Tissue Inhibitors of Metallo Proteases) (case study C);
- a Japanese study on occupational lifting (case study B);
- subsequent research on the effect of antiphospholipid antibodies on *in vitro* fertilisation (case study K); and
- a significant, possibly paradigm-changing, shift in the use of biologicals as therapeutic agents (case study L).

Finally, it is worth noting the development of reagents (case study C) and a mouse model (case study D) as further examples of research payback. In the former, although records were not kept, the PI estimated that the reagents produced during the grant were sent to over 100 other research groups.

In summary, the 16 case studies illustrate a comprehensive and diverse range of research targeting and capacity building paybacks.

3.1.3 Informing policy and product development

As with the previous two categories, we identified a range of different types of payback that inform either policy development or product development. These include: being cited on systematic reviews; being cited on clinical guidelines and other forms of policy guidance; and the development of specific clinical tests.

Four papers that were attributable to case studies A and B were cited subsequently in systematic reviews. Moreover, in both of these examples the original papers were independently “quality assured” by the systematic review authors and all received a high assessment. Seven papers from the case studies were cited on clinical guidelines relating to arthritis. Case study P had the highest level of citation, with three papers cited on arthritis-related guidelines, and one paper cited by the non-arthritis-related guideline. Two papers from case study L, the development of anti-TNF drugs, were cited and both case studies E and K had one paper that was cited by arthritis-related guidelines.

More directly attributable payback may occur when the research supports clinical guidelines. This was most apparent for case study K, where the findings were an important element in the British Royal College of Obstetricians and Gynaecologist guidelines recommending a combination therapy of aspirin plus heparin to prevent recurrent miscarriages for women who have antiphospholipid syndrome (APS).¹⁰ Similarly, national guidelines for physiotherapy treatment now entail proprioceptive exercise as a component of anterior cruciate ligament deficiency (ACLD)¹¹ management, although it is not entirely clear how much of this can be directly attributable to case study J and how much of it reflects a body of research to which case study J contributed. Less directly, papers attributable to case study B were cited in a Dutch clinical guideline; similarly for case study E on a European guideline (which was authored by the PI). Finally, epidemiological data arising from the Norfolk Arthritis Register (case study P) has been widely cited in the introduction to a number of guidelines.

An especially interesting example of payback arose from case study B. This grant looked at the role of occupational activity and hip osteoarthritis (hip OA)¹² and concluded that agricultural workers were at an increased risk of hip OA. This work was cited in the Industrial Injuries Advisory Council assessment of whether hip OA in farmers should be a prescribed disease (and thus receive Industrial Injuries Benefit); the Council’s subsequent advice to the Secretary of State was that this was justified. In addition to the actual paper, one of the lead PIs gave a presentation to the Council.

In terms of informing product development, two case studies identified potential drug targets. Case study H indirectly informed the development of a class of drugs used outside the arthritis field to treat deep vein thrombosis. Case study L tested the hypothesis that rheumatoid arthritis (RA)¹³ could be treated by using targeted TNF alpha inhibitors.

Finally, two case studies – F and G – identified specific diagnostic tests, although these were for very rare conditions. The first was for specific mutations in collagen X and used

¹⁰ This is a disorder of the immune system characterised by excessive blood clotting and/or pregnancy losses, and is associated with typical laboratory abnormalities including persistently elevated levels of antibodies directed against phospholipids (part of the cell’s membrane) or their associated plasma proteins.

¹¹ The anterior ligament crosses the front of the knee joint. It keeps the knee from sliding backwards and forwards and therefore plays a crucial role in activities such as running and descending stairs. Injury and tearing of the anterior ligament is associated with joint instability and OA.

¹² Hip OA of the hip causes the cartilage between the pelvis and the thigh bone to wear away, leading to pain and limitation of hip motion.

¹³ A chronic inflammatory disease in which the body’s immune system attacks the joints, causing pain, joint destruction, disability and premature mortality.

to test for chondrodysplasia type Schmid.¹⁴ The second enabled the measurement of antibodies to C1q in patients with systemic lupus erythematosus (SLE).¹⁵

3.1.4 Health and health sector benefits

Three grants have had significant health and health sector benefits. Case study K concluded that a combined treatment of aspirin and heparin for women with APS – which affects about 15 in 10,000 women – reduces the risk of recurrent miscarriages and consequently means that these women have a far higher chance of a successful pregnancy. The findings have been strongly promoted.

Case study L showed that in approximately 70% of patients with RA – which affects about 1 in 100 people – anti-TNF treatment leads to significant health improvements. Nevertheless, to date, uptake has been limited due to the lack of resources to fund the treatment at £10,000 per patient year. For example, in the UK the uptake is approximately 2% of patients; in Scandinavia, approximately 6%; and in the US, approximately 15%. Nevertheless, this still translates to over 8,000 people in the UK, and over a quarter of a million in the US, who are experiencing health improvement. Despite this, a recent study concluded that during the first year of treatment direct costs were reduced by 40%, while indirect costs did not change substantially. The estimated cost per quality adjusted life year gained was estimated at €50,000 (approximately £33,000) (Kobelt and others 2004).

Although case study P concerned the epidemiology and prevalence of RA, it was seen to be providing health benefits by its identification of other risk factors for RA patients – such as heart disease – that were affecting patient care.

For the more basic research is it still too early to judge what the total benefits may be; for example, regarding case study C, there are active drug development programmes that grew out of the research. In other cases such as case studies E and F, improved understanding of the disease processes of OA may lead to future improvements in treatment; similar benefits may come out of case study H for RA patients. Finally, there are case studies where the technology and techniques developed laid the groundwork for current basic research that eventually may lead to patient benefit: case study O has led onto work in pharmacogenomics that may benefit RA patients.

3.1.5 Broader economic benefits

In our case studies, there was no quantifiable evidence of broader economic returns arising from the arc-sponsored research. For cases A, B, J and K there are unquantified returns in the reduction of days off work and the value of production gained from having a more healthy workforce. In addition, for case study L, there are economic returns to the three companies that have licensed drugs in the UK, although these returns need to be considered in a global context given the multinational structure of the companies.

¹⁴ A very rare disorder characterised by abnormal development of skeletal tissue.

¹⁵ A poorly understood chronic inflammatory disorder in which the immune system is overactive, producing abnormal antibodies that react with the patient's own tissues. The disorder may cause skin rashes, arthritis, seizures, psychiatric illness and damage to internal organs.

3.2 Analysis of the payback model

In this section we describe the case studies by the seven stages and two interfaces of the arc Payback Model, based on the summaries provided in Table 3.3. As some of this analysis overlaps with the cataloguing of payback categories (as is the case for knowledge production and primary outputs), we focus on those aspects of the framework that have not been discussed previously. Finally, we look at the common “payback pathways” by which research is translated from output to outcome.

3.2.1 Stage 0: topic/issue identification

As would be expected given the response mode nature of arc funding, all the case studies arose from ideas initiated by the PIs, sometimes in collaboration with their supervisors or research mentors. In about half of the case studies, the proposed work built on previous arc funding. The arc-funded work that was built upon included a project grant and senior research fellowship for case study C, a selection of previous project grants for case study E and a junior training fellowship and project grant for case study G. In addition, the work preceding case study F and the four institute grants (L, N, O and P) was arc-supported. Finally, it is worth noting that four of the proposals were the applicants’ first successful applications to arc (case studies B, D, I, K), and two were re-applications (case studies A and I).

3.2.2 Interface A: project specification and selection (peer review)

Within our 16 case studies, a number of different models of peer review were used (detailed in Box 3.1). Some of the peer-review systems incorporated quantitative ratings, whereas others relied entirely on qualitative comments. Where quantitative rating was used, referees were asked to score grants on a three-point scale against seven criteria. For the six project and two programme grants for which this system was applied, the distribution of referees’ assessments is shown in Figure 3.1. Not surprisingly, given that we were looking at awarded grants, each of the applications was rated highly by the majority of referees on all but one of the seven criteria (the exception being “feasibility within time proposed”). Perhaps of more interest was the level of disagreement between referees and the relatively large minority of referees rating the applications as medium or low. Disagreement was greatest for “originality of project” and “feasibility within time proposed” and smallest for “potential practical value” and “standing of applicants in the field”.¹⁶ In terms of the total number of assessments, 55% of ratings were high, compared to 45% which were medium or low.

Mode of funding

arc’s system of peer review has evolved significantly since the early 1990s. There has been an increased use of international referees and steps have been taken to improve representation of relevant disciplines on review committees.

Project grants:

External reviewers were asked to rate the proposals against a number of specific criteria, and also to provide qualitative comments. Sometimes these comments were fed back to the applicants, but only at the discretion of the award committee.

¹⁶ Disagreement is crudely measured as the difference between the number of referees rating high *versus* medium for each of the seven criteria.

<i>Programme grants:</i>	Although the programme grants awarded during the selection window were not arc's first programme grants, there was no separate programme grants committee. Applications were allowed at any time in an ad hoc fashion, rather than in open competition. Initial application was by letter of intent: this letter was reviewed by the project grants committee, and comments might be fed back to the applicant. Successful applicants were asked then for a full application which was assessed by external reviewers, who on occasion used the same scoring system as that for project grants. Finally, a site visit was carried out before the decision to award was taken.
<i>Fellowships:</i>	Initially, applications were assessed by external reviewers. These reviewers' comments were considered by the award committee, along with the comments of the applicant's personal referees. Shortlisted candidates were invited then for interview before the final decision to award was made.
<i>Institutes:</i>	Towards the end of the 1980s discussions began about formalising the institute review process. During our selection window this change was implemented, moving to a quinquennial (five-yearly) review system. In addition, during the selection window institute funding was consolidated into single core grants and the institutes were prevented from applying for additional funding from arc in the form of project, programme or fellowship grants. The review processes applying to the institute-based grants are discussed in more detail in the relevant case studies.

Box 3.1: **arc** peer-review systems in use between 1991 and 1994

Table 3.3: Payback model populated for 16 case studies

Stage 0 Case Study	Opportunity identification	Peer review	Interface A	Stage 1	Stage 2 Process	Stage 3 Primary outputs	Interface B Dissemination	Stage 4 Secondary outputs	Stage 5 Adoption	Stage 6 Final outcomes
A	The proposal was developed against the background of an apparently increasing burden of back pain. The PI and colleagues had developed a general exercise programme for patients with chronic low back pain. The proposal was to assess the effectiveness in community settings of the exercise programme.	One of four referees gave the project "strong support"; the other three "possible support". The application was not accepted, but detailed queries were fed back to the applicant. On resubmission all referees said the application had been improved.	An award of c. £150,000 over 3 years. Additional funding to support aspects of the project came from the National Back Pain in Association and the Northern and Yorkshire NHS Executive. The research was conducted at the Centre for Health Economics at York University which was able to offer academic support and liaison.	There was a delay in recruiting patients to the trial, partly due to the illness of one of the original research physiotherapists employed on the project. Once the initial difficulties were addressed, the cost-effective. The project provided research training and read peer-reviewed informed the evaluation of the "Back to Fitness" programme and has been highly rated in several external reviews of work in this field, for the methods that were used.	There has been a general policy move towards more active management of back pain. National guidelines on this pre-date the publications from this work, although it has been included in a task group report. At a local level it is influencing policy in various places.	The PI gave many presentations at academic and professional meetings, quite often featuring the findings from the study. In addition, the project was disseminated through an article in a widely read journal and informed the evaluation of the "Back to Fitness" programme through a series of study days run by the PI for physiotherapists.	The PI gave many presentations at academic and professional meetings, quite often featuring the findings from the study. In addition, the project was disseminated through an article in a widely read journal and informed the evaluation of the "Back to Fitness" programme through a series of study days run by the PI for physiotherapists.	The study showed that there could be some gains in terms of benefits to the health sector, such as health gains, cost savings, and fewer working days lost. Without knowing the extent of adoption of the "Back to Fitness" programme it is difficult to make a full assessment of these final outcomes.	The study showed that there could be some gains in terms of benefits to the health sector, such as health gains, cost savings, and fewer working days lost. Without knowing the extent of adoption of the "Back to Fitness" programme it is difficult to make a full assessment of these final outcomes.	
B	The proposal arose from three observations made by the research team. First was the higher prevalence of OA in farmers. Second was the development of a definition of radiological hip OA. Third was the training and methodological development in a study of knee OA.	Three out of four referees gave "strong support" to the study. The fourth (an epidemiologist) gave "possible support", citing concerns regarding the design of the study.	An award of c. £80,000 over 2 years.	There was a delay in recruiting two research nurses for the study.	There was a delay in recruiting 6 papers, receiving a total of 5 citations per year. The research concluded that an exercise class is more effective than traditional GP management and is more cost-effective. The project generally ran smoothly and has been highly rated in several external reviews of work in this field, for the methods that were used.	Three papers were cited in three systematic reviews, each receiving a high-quality assessment. One paper was informed by the Industrial Injuries Advisory Council (IAC) Assessment on whether hip OA in farmers was a prescribed disease.	Three papers were cited in three systematic reviews, each receiving a high-quality assessment. One paper was informed by the Industrial Injuries Advisory Council (IAC) Assessment on whether hip OA in farmers was a prescribed disease.	The adoption of the IAC recommendation will result in Industrial Injuries Disability Benefit to farmers with hip OA.	Farmers with hip OA will receive Industrial Injuries Disability Benefit.	
C	The PI was an established researcher working on MMPs, having been awarded an arc proposal and a five-year fellowship. The project took advantage of the discovery of TIMP-2 in 1989 and the possibility of comparing the properties, mechanism and distribution of the two inhibitors.	The proposal was reviewed by three referees. Two referees "strongly supported" the proposal, and the third noted that it was an "excellent grant application".	An award of £150,000 over three years. The included a 15% "establishment cost" for overheads at Strangeways Research Laboratory. In addition, the PI held a complex portfolio of grants from arc and the other funders. One of the key inputs was the expertise of reagents and techniques developed in the PI's laboratory.	First two strands of the work proceeded as in the grant application with no major changes of direction. The final strand of work, looking at a portfolio of grants from arc and the other funders. One of the key inputs was the expertise of reagents and techniques developed in the PI's laboratory.	Range of unspecified activities, as part of normal academic discourse.	Celitech was interested in recombinant TIMPs as a therapeutic. The PI's contacts were key in setting up this partnership. Although much of the project was successful when a knockout mouse was produced to work in the unit where the work was conducted.	Range of unspecified activities, as part of normal academic discourse.	Because none of the drug development programmes that have been produced no final outcomes in terms of health gain can be attributed to this grant.	As no registered products have been produced no final outcomes in terms of health gain can be attributed to this grant.	
D	This was the first research grant that the PI had applied for. The proposal was a continuation of his work and was influenced by (1) the need for animal models given the acute shortage of samples from patients with reactive arthritis; and (b) the development of a transgenic rat which was associated with reactive arthritis.	The proposal was reviewed by three referees. One indicated strong support, the other indicated possible support and the third fell between the two. One referee suggested that the candidate should apply for a fellowship.	An award of c. £100,000 over 3 years. The grant was funded by arc from a donation from Glaxo. An important, non-financial, input was the hanging drop assay technique that allowed for measuring the proliferation of small numbers of T cells. This had been developed in the laboratory where the PI worked prior to the grant.	The PI spent 6 months in the US working on transgenic rats. The work done on the grant as the work done on the grant as they do not directly concern chlamydial infection or reactive arthritis. The PI's career has continued to advance, but outside the field of arthritis.	Range of unspecified activities, as part of normal academic discourse.	The grant produced three outputs which focused on infectious diseases and do not have a direct relevance to arthritis. The PI developed a mouse model of chlamydial infection. The work defined a peptide that was a promising vaccine target. The identification of the chlamydia peptide used a method of finding vaccine targets.	Range of unspecified activities, as part of normal academic discourse.	No evidence of any final outcomes.	No evidence of any final outcomes.	

Table 3.3: Payback model populated for 16 case studies (continued)

Case Study	Stage 0 Opportunity identification	Interface A Peer review	Stage 1 Inputs	Stage 2 Process	Stage 3 Outputs	Interface B Dissemination	Stage 4 Secondary outputs	Stage 5 Adoption	Stage 6 Final outcomes
E	The research built on a previous arc-supported programme that enabled the team to recruit a patient cohort and associated infrastructure. It focused on knee OA as it was the most painful condition in patient cohort.	Assessment on the basis of a letter of intent, proposal and a site visit. The letter of intent clinched significant sub-set of referees who all gave it their "strong support". The site visiting team felt that the basic science element of the programme was not in the front-line of innovations'.	An award of c. £500,000 from the arc. The multidisciplinary team were pursued, with a focus on: OA; subchondral bone activity and the interaction between the cartilage and the bone at the cellular level. The mid-term postdoctoral presentations established several of the junior clinicians have become national and international OA experts. The PI became a leader in HSR.	Half of the original hypotheses were pursued, with a focus on: OA; subchondral bone activity and the interaction between the cartilage and the bone at the cellular level. The mid-term postdoctoral presentations established several of the junior clinicians have become national and international OA experts. The PI became a leader in HSR.	44 papers, receiving a total of 73 citations per year. One referee praised the PI's ability to communicate his research effectively. The PI gave a contributed to two books. Two presentations at international conferences, while the team hosted an international symposium on OA.	In the programme renewal, a referee praised the PI's ability to communicate his research effectively. The PI gave a contributed to two books. Two presentations at international conferences, while the team hosted an international symposium on OA.	The medical taping technique, which was developed on the programme, was cited in a European League Against Rheumatism clinical guideline that was authored by the PI.	The interviewees believed that there has been very little change in OA treatment, even in the local region or host institution.	From a limited number of patients benefitting from the biomechanical techniques that were developed, patients health had not improved as a result of the programme.
F	Built on the PI's previous work to the identification of collagen X at sites of endochondral ossification and the subsequent cloning of human collagen X.	This was one of the first arc programme grants awarded to a non-rheumatologist. The review process was changing and took over 6 months. Decision was based on 3 referees' comments and a site visit. Both the referees and site visit were worried that the programme was too ambitious and thus only 3 of the 4 scientific aims were supported.	An award of c. £450,000 over five years. At the time, the overall funding represented about 20% of his funding.	Three years into the programme grant, the team had fulfilled the objectives of the applications. In addition, they had identified two mutations that caused chondrodyplasia type Schmidt, a very rare skeletal dysplasia with a prevalence of about 1 in 10 million people.	23 papers, receiving a total of 32 citations per year. The research showed that collagen X was a definite marker for endochondral ossification and identified specific mutations in collagen X that cause chondrodyplasia type Schmidt. The two research assistants obtained their PhDs because of the programme.	Range of unspecified activities, as part of normal academic discourse.	The research was picked up by another academic group who used the identification of specific mutations in collagen X to develop a diagnostic test for chondrodyplasia type Schmidt. This test is currently run as a clinical service for the European skeletal dysplasia network.	The test is in use, but at patients' request.	Confirmation of type Schmidt chondrodyplasia – as is given by the diagnostic – influences the management of care, and specifically prevents surgical intervention that may be applicable to other forms of chondrodyplasia.
G	The fellow previously was awarded an arc junior training fellowship and project grant that characterised the mechanisms involved in the processing of ICs. This led to the hypothesis that the disease mechanism in SLE was associated with abnormal immune processing mechanisms.	Eight referees reviewed the proposal and were overwhelmingly supportive.	An award of c. £400,000 over five years from arc. One referee noted that importance of the longstanding collaboration with the Departments of Medicine and Radiology at Hammersmith Hospital. In addition, the PI stressed the intellectual stimulation arising from working at the Hammersmith Hospital.	No major changes to the course of action specified in the proposal. After 3 years, the PI was appointed senior lecturer and therefore transferred the fellowship funds to support another senior research assistant.	32 papers, receiving a total of 56 citations per year and 1 1C1q antibodies in SLE. The importance of anti-C1q antibodies in patients with SLE was also demonstrated.	Preliminary findings of the research were presented at the British Society of Rheumatology meeting. The PI also won the Michael Mason Prize, involving making a presentation to the British Society of Rheumatology. In addition, there was the range of unspecified activities, as part of normal academic discourse.	The development of a diagnostic test that enabled the measuring of antibodies to C1q in SLE patients.	The test is still used today at Royal Postgraduate Medical School, at Hammersmith Hospital.	No evidence of any final outcomes.
H	The application followed a previous proposal for a senior research fellowship that had failed because it was deemed too ambitious. Learning from this experience, the PI proposed a more conventional driven descriptive study.	The four referees noted the strength of the candidate and his working location, but expressed some concerns about its theoretical basis and that it was a non-hypothesis-driven descriptive study.	An award of c. £250,000 over 5 years (supported by a personal bequest). This supported the PI's salary, plus a technical assistant (TA).	Through the fellowship, the PI built an active research team, pursuing further work on ICs. Main findings: the relationship between T cell rheumatoid process is not to be grossly abnormal; the differentiation, apoptosis, homeostasis and cytokine regulation. The reduced tenreure interaction between T cells and stromal cells. RGD peptides cross the cell membrane of T cells and trigger an apoptosis-signalling pathway causing cell death.	Range of unspecified activities, as part of normal academic discourse.	The observation that RGD peptides were triggering apoptosis led to the development of KGD peptides, a class of drugs used to treat deep vein thrombosis, but not useful for arthritis.	No evidence of adoption.	No evidence of any final outcomes.	

Table 3.3: Payback model populated for 16 case studies (continued)

Stage 0 Opportunity identification	Interface A Peer review	Stage 1 Inputs	Stage 2 Process	Stage 3 Primary outputs	Stage 4 Secondary outputs	Interface B Dissemination	Stage 5 Adoption	Stage 6 Final outcomes
J	The proposal was a research application. The three referees agreed that the proposal was application, as the supervisor was able to support the PI from 'soft money' funded by industrial collaborators.	An award of c. £55,000 over 2 years. During the process of application, the three referees agreed that the proposal was a good training exercise, but had the supervisor was able to support the PI from 'soft money' funded by industrial collaborators.	A delay of one year occurred in the construction of a device to assess quadriceps muscle strength. There were also difficulties with patient recruitment, but these had been sorted out by the time the equipment was completed. As the study did not deliver what was expected PI also finished his MD arc did not award an extension and later became a consultant in rheumatology.	11 papers, receiving a total of 32 citations per year and 1 PhD. The research hypothesis was not addressed in the timeline of the fellowship, but the research contributed to the number of methodological developments including a grading system to measure OA. PI also finished his MD arc did not award an extension and later became a consultant in rheumatology.	Range of unspecified activities, as part of normal academic discourse.	Peripheral to the fellowship, the PI used his knowledge in developing a grading system to help validate the 'GALS' grading system. The lessons applied to a grant that tested the same hypothesis which subsequently informed a US RCT assessing the effect of exercise on people with OA.	Although not demonstrated as part of the fellowship, the PI used his knowledge in developing a grading system to help validate the 'GALS' grading system. The lessons applied to a subsequent project grant that tested the same hypothesis which subsequently informed a US RCT assessing the effect of exercise on people with OA.	No evidence of any final outcomes. However, based on the adoption of exercises for OA patients, health benefits may be expected.
J	The PI responded to an advertisement for a physiotherapy research in proprioception lead by a surgeon, who had identified a US study on proprioception as an important component of rehabilitation. The surgeon wanted to recruit someone with a background in physiotherapy to evaluate this observation.	An award of c. £45,000 over 2 years. In addition, access to on-site engineering facilities and equipment was essential to the research.	Three referees reviewed the proposal and recommended the award subject to confirmation of surgical collaboration, evidence of ethical approval and affirmed support of the head of department.	13 papers, receiving a total of 10 citations per year. These showed: (1) that proprioception plays an important role in ACLD deficiency, and (2) a rehabilitation programme to enhance proprioception was more effective than traditional muscle-strengthening exercise. The PI was awarded a PhD for the work.	Through the production of brief summaries for user groups and study days and training at local clinics.	A novel technique in measuring proprioception; capacity building through knowledge transfer nationally and internationally.	The PI's research effected a change in physiotherapy ACLD/ACLR management approaches, one which now includes proprioceptive exercises in national clinical guidelines and care pathways.	The work influenced the change in management of ACLD/ACLR patients. Physiotherapy practice now includes proprioceptive exercises.
K	During the 1980s rheumatologists observed that those referees, although they viewed the science as good, only thought the topic too far removed from mainstream rheumatologic research to merit funding by arc. Two of the referees thought that funding should be substantially reduced once a marker for APS, was treated experimentally with aspirin and sum of c. £100,000.	An award of c. £45,000 over 2 years. The substantial cut was on the grounds that a research persistently positive tests for the project could coordinate the antiphospholipid antibodies were randomised to receive low-dose aspirin or low-dose aspirin plus heparin. The PI disputed this and after some negotiation with arc funding for 34th week of gestation. All women took low-dose aspirin from the time of a positive pregnancy test and were randomised once foetal ultrasound. An 8-month extension added c. £20,000 to the total budget.	86 women with a history of recurrent miscarriage and 20 citations per year and 1 PhD. The main paper concluded that treatment with aspirin and heparin leads to a significantly higher rate of live births in women with a history of recurrent miscarriage associated with phospholipid antibodies than that achieved with aspirin alone. The trial resulted in an international profile for the research team and the research register got a commendation for his MD thesis on the project.	In addition to normal academic discourse, the PI provided lay information in publications and meetings of patient groups. This included a book on miscarriage, and radio discussions. The PI's work was featured in a Channel 4 documentary.	The British Royal College of Obstetricians and Gynaecologists guideline, along with US, Dutch and Australian Guidelines, all recommend a combination of aspirin plus heparin and all cite the trial as the evidence for this recommendation.	Assuming that these authorities' guidelines are being followed (and this is the perception) then the combined treatment of aspirin and heparin has been widely adopted.	The trial led to the conclusion that a combined treatment increased the chances of a live birth for women who had tested positive for antiphospholipid antibodies. A substantial number of these women would not have successfully completed a pregnancy without this treatment.	
L	From an interest in the role of cytokines in RA, the research team established which cytokines were present in RA tissue and then investigated whether they could be blocked in therapy. By 1992 success in blocking TNF with antibodies had been shown in animal models, in order to test whether this effect could be obtained in patients a small clinical trial using anti-human TNF antibody was proposed.	The funding for the division was awarded without revision as part of the £2.7 million, 5-year institute core grant. It is no longer possible to estimate how much of this was for the proposed trial. In addition to arc funding, support of over \$1 million was obtained from a US industrial partner. At the time there was little interest from UK companies.	An initial small trial of 20 patients was followed by a large out-centre double blind trial. There were a number of delays in persuading the industrial collaborator to support the second trial due to a shortage of funds and concerns about the primary use of anti-TNF.	In approximately 70% of patients, anti-TNF treatment leads to significant improvements. A recent study concluded that during the first year of treatment, direct costs were reduced by 40%, while indirect costs did not change substantially. Estimated costs per quality adjusted life year gained amounted to €50,000.	The highest uptake of treatment has been in the US, where an estimated 15% of RA patients benefit. In Europe adoption is highest in Scandinavia (6-7%), followed by Spain and the Netherlands. Uptake in the UK is low (2%).	The research of the team led to the development of three drugs - infliximab, etanercept and adalimumab - which have been licensed in the UK. Anti-TNF treatment has been included in the guidelines of the national news item of the national news which resulted in around 5,000 patients in the UK and the Netherlands. Phone calls being made to the institute, adopted by international consensus groups in Europe and the US.	The research of the team led to the development of three drugs - infliximab, etanercept and adalimumab - which have been licensed in the UK. Anti-TNF treatment has been included in the guidelines of the national news item of the national news which resulted in around 5,000 patients in the UK and the Netherlands. Phone calls being made to the institute, adopted by international consensus groups in Europe and the US.	

Table 3.3: Payback model populated for 16 case studies (continued)

Case study	Stage 0 Opportunity identification	Interface A Peer review	Stage 1 Inputs	Stage 2 Process	Stage 3 Primary outputs	Interface B Dissemination	Stage 4 Secondary outputs	Stage 5 Adoption	Stage 6 Final outcomes
N	The work resulted from the PI's increased interest in RA and the development of genome-wide screening for other autoimmune diseases.	Four referees reviewed the proposal and were supportive of the project, but voiced concerns over the ease of recruiting suitable families for the extension of core facilities already established to map genes.	An award of c. £500,000 over 3 years was made. £230,000 of this was capital expenditure for the extension of core facilities, already established to map genes.	During the grant a number of issues regarding the ownership of data arose between collaborating centres. In addition, the PI became more engaged in the pursuit of 1996, 170 sibling pair families had been identified, confirming the genetic basis of ankylosing spondylitis (AS), which linkage to HLA.	15 papers, receiving a total of 20 citations per year. The research identified a number of specific genetic regions and AS.	In addition to normal academic media attention for his work on HLA, but the evidence is not statistically powerful, and therefore this research did not lead to any commercial leads or applications. However, it stimulated research in AS, which continues to date.	There is evidence of linkage for a number of genes other than HLA, but the evidence is not statistically powerful, and therefore this research did not lead to any commercial leads or applications. However, it stimulated research in AS, which continues to date.	There have been no opportunities for adoption.	No evidence of any final outcomes.
N	The PI's interest in proteoglycans dated back to his PhD and subsequent work. As part of the work plan for the programme grant was reduced from 5 to 3 years with extension conditional upon the PI demonstrating the capacity to map and identify genetic linkages.	The proposal submitted by the institute's biochemistry division was awarded without revision as part of the £2.7 million 5-year institute grant. Due to the accounting practice at the time it is not possible to estimate how much of this was for the proposed projects.	In interview the PI pointed out that (with the benefit of hindsight) the team were hindered by the absence of tools which now would enable them to undertake the proposed work.	16 papers, receiving a total of 75 citations per year and 3 PhDs. One reviewer of grant renewal noted that the work led to few journal publications in first class journals. However, the PI argued that the work had had a strategic influence on the field. The PI's refocused his research interests when he took up a chair in Manchester in 1994.	Range of unspecified activities, as part of normal academic discourse.	Work on the basic structure has thrown light on the structure and shape of complex cellular molecules involved in OA, but has not led to any applications.	There have been no opportunity for adoption.	No evidence of any final outcomes.	
O	The PRF fellowship developed from the establishment of a UK2 referees and 2 assessors. The panel found it difficult to assess the application due to concerns raised about the number of patient samples needed for linkage studies. The panel responded by stating there was a delicate balance between their desire to give an award, and these concerns. Accordingly, the panel awarded a 2-year fellowship with a suggestion that progress would be assessed at 18 months.	An award of c. £180,000 was made. In addition, the collaborating centre in genome studies was essential to the study.	During the grant a number of issues arose regarding competition between the institute and its collaborating centre. The basis of this was the desire of both to set up a cohort of patients from whom to collect DNA. arc originally decided that the two groups would have to work together, but this was difficult.	17 papers, receiving a total of 21 citations per year and 1 PhD. The core publications describe the identification of dinucleotide repeat polymorphisms as the alpha-1-antitrypsin locus and give a summary of the first 100 multi-case RA families with affected sibling pairs. The fellowship benefited the PI's career but also the RA genetics capacity at the institute.	Range of unspecified activities, as part of normal academic discourse.	The institute did not have any industrial collaborations in the period under examination, however, as its reputation has grown in RA genetics pharmaceutical companies have become more interested in its activities.	There has been no opportunity for adoption.	No evidence of any final outcomes.	
P	The idea for an arthritis register was stimulated by the success of population-based cancer registers created by the PI and colleagues in the field. The idea was brought to the Institute with the appointment of the PI as director.	No assessment through a formal process of peer review, except as part of the PIs application to become Institute director. However, it is unclear whether the appointment committee was able fully to evaluate the research. As institute director, the PI established a formal peer-review process and asked arc to set up a site visit. The visiting team concluded that NOAR had "major potential".	Upon completion of a pilot study, the register was launched at a press conference in 1990. All 276 GPs in 76 practices in the Norwich Health District were assured of long-term institute funding was deemed critical by the PI; registrants need several years to develop both infrastructure and human capital before significant scientific outputs emerge.	27 papers, receiving a total of 40 citations per year. NOAR has provided groundbreaking work in understanding how common RA is and the factors that are associated with it. The <i>Arthritis Today</i> , arc's quarterly lead-co-ordinators and lead publication.	Interviewees noted that NOAR has increased research capacity locally in Norfolk. Indirectly this may have influenced the establishment of a new medical school. The Danish have set up a similar registry based on NOAR. The incidence and prevalence figures from NOAR are widely cited in clinical guidelines.	Recent data from NOAR on the benefit of early treatment has become part of the evidence base supporting early referral and treatment. Early referral of RA patients is one of the two quality standards set for the Royal College of Physicians.	No evidence of any final outcomes.		

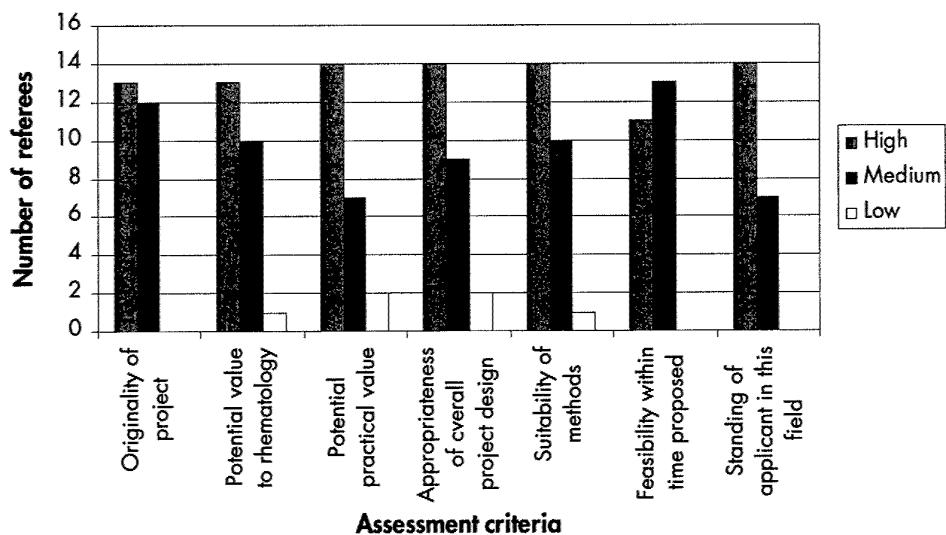


Figure 3.1: Peer-review assessment of six project grants and two programme grants

The preceding quantitative analysis of seven applications illustrates a degree of ambiguity in the referees' assessments. This was confirmed also by a qualitative review of the referees' comments that were provided for all 16 case studies. Moreover, it is not clear as to the degree to which the referees' comments added (scientific) value. In the case of the two resubmissions (A and I) the PIs commented that the referees' comments had improved the projects; however, in the remaining 14 cases only three PIs recalled positive changes in light of the referees' comments and in two of these cases the improvements were slight.

In addition to judging the science, referees sometimes made recommendations to reduce the proposed budget. In two cases (M and O) the grant was reduced with extension conditional on progress; in the first case this was not granted, and in the second the grant was subsumed into an institute grant, so the issue of renewal did not arise. In a further three cases (E, F and K) the referees suggested a budget cut. In two of these cases (F and K) the award committee made cuts; in both cases the PIs found themselves without sufficient funds and subsequently requested additional money from arc. In case study A the health economic/cost benefit strand – around 7% of the total cost – was cut from the project during the review process, but was funded subsequently from another source.

As a final observation: for the grants resulting in the largest levels of payback (K and L), if the referees' comments had been taken at face value rather than being overruled by the award committee, it is unlikely that they would have been funded.

3.2.3 Stage 1: inputs to research

In total our 16 case study PIs were awarded grants worth over £3 million (this excludes two of the institute projects for which it is only possible to estimate an approximate value). For comparison, in 1990 and 1994 arc awarded just over £50 million. In addition, and as noted in Table 3.3, there were a range of other inputs, including:

financial support from other not-for-profit and commercial funders (case studies A, C and I);

the availability of reagents, techniques, tissue samples, etc (case studies C and E); and

access to specific pieces of equipments, facilities and space (case studies E, J, M and P).

Finally, as noted above, the majority of the grants were a continuation of previous research, thus one of the major inputs was the pre-existing tacit knowledge of the PIs. It was also notable that many grants benefited from research facilities or the assistance of individuals which were never charged to the grant: the use of radiology facilities, for example. It is likely that nowadays such facilities and staff time would be charged to the researcher, making the research appear more expensive.

3.2.4 Stage 2: research process

The PIs provided a rich and valuable source of information about all stages of the research. To reduce the chances of biased recall, generally we tried to corroborate the PIs' recollections with information from other sources. However, when dealing with issues surrounding the research process we had to rely far more heavily on the PIs' statements, or those by members of their research group, as corroborating sources were much more difficult to identify.

Two broad themes emerged when examining the research process: issues of personnel and unforeseen technical difficulties. First, personnel issues encompassed the recruitment and retention of staff on projects (including, in some cases, the PI). For example, in case studies A and B the PIs reported difficulty in recruiting research staff to the two project grants and this resulted in a delay to the work beginning. For case study J, the PI first took a sabbatical and then was offered and accepted a part-time senior management position in his institute; both again had the effect of delaying the project. Another major personnel issue was the turnover of junior research staff. In the case studies that we examined **arc** always approved cost-neutral changes, where money was moved from supporting one junior researcher who was leaving to their replacement. **arc** also showed flexibility in allowing the promotion of junior research staff, where the increase in salary was balanced against a decrease in the length of support.

Second, there were difficulties in performing the actual research that had been specified in the proposal. In case study C, a strand of work proved harder than anticipated (although the work was successfully carried out sometime later). Similarly, case study D ended up by focusing on two minor elements of the project grant proposal, due to problems with the main strands of work (mostly concerning the availability of sample material). The PI for case study D was a first-time applicant to **arc**¹⁷ and he subsequently moved into a non-arthritis field of research following a further, failed **arc** fellowship application. For case study I, there were difficulties in recruiting patients to the study, but a limiting factor was a delay caused by problems in constructing a specific piece of equipment. In case study L, difficulties in identifying and persuading an industrial partner to collaborate on the trial delayed the project. Finally, in case studies M and O, related issues regarding the ownership of data between collaborating centres and competition between two centres in establishing a patient cohort characterised both projects.

¹⁷ Although he was previously supported as a postdoctoral researcher on an **arc** grant held by his laboratory head.

3.2.5 Stage 3: primary outputs from research

As discussed in more depth in section 3.1.1 on knowledge production, 302 papers receiving an average of 975 citations per year were published as a result of the 16 case study grants awarded by arc (Table 3.1). In addition (and as discussed in section 3.1.2 on research targeting and capacity building) seven people were awarded either a PhD or MD as a result of the research that was undertaken on the 16 grants.

3.2.6 Interface B: dissemination

All the PIs confirmed that they had disseminated their research through conference papers, seminars and other forms of academic discourse, although generally they had not kept detailed records of such events. In addition, the one of the AHP case studies (A) had used study days to disseminate the findings of their research to their peers – although arc provided no support for these activities. Two other PIs identified forms of public engagement, including a lay presentation to the executors of the estate that supported his research (case study H), speaking at patient group meetings and participating in a Channel 4 documentary (case study K). Finally, the successful results of the anti-TNF treatment were released at a press conference and achieved widespread coverage (case study L).

3.2.7 Stage 4: secondary outputs – policymaking and product development

There is a strong correlation between this stage and the payback category on informing policy and product development. As discussed in section 3.1.3, we identified a range of secondary outputs including:

- being cited in systematic reviews (case studies A and B);
- being cited in clinical guidelines and other forms of policy guidance (case studies B, E, J and K);
- the identification of drug targets (case studies C, H and L); and
- the development of specific clinical tests (case studies F and G).

3.2.8 Stage 5: adoption – by practitioners and public

In three case studies there was evidence that clinical management had changed due to the uptake or adoption of the tested intervention. In case study B there was some local evidence that physiotherapists had used the “Back to Fitness” programme. Similarly, it was claimed that it is now common for clinicians to recommend physical training for people with OA (case study B). For case study L, the highest uptake of anti-TNF treatment has been in the US, where an estimated 15% of people with RA benefit from it. In Europe, adoption is the highest in Scandinavia (6–7%), followed by Spain and the Netherlands. In the UK uptake is low, at around 2%.

3.2.9 Stage 6: final outcomes

The final outcomes of the arc Payback Model were summarised in sections 3.1.4 and 3.1.5, describing the health and health sector benefits, and the broader economic benefits payback categories. These included:

- the reduced risk of recurrent miscarriages for women with APS (case study K);
- improved treatment for people who suffer from RA (case study L); and
- in-depth understanding of the prevalence, characteristics and causes of arthritic disease (case study P).

In one case the diagnostic test developed for chondroplasia type Schmidt leads to avoidance of surgical intervention as this has been shown to make the symptoms worse; however, this disorder is extremely uncommon so the overall impact is likely to be slight.

3.3 Cross-case analysis

In order to assess the strengths and weakness of different modes of funding and types of research, and to examine the impact that a researcher's track record has on translation, we undertook a series of comparative analyses of the case studies. The cross-case analyses were based on a qualitative examination of common types of payback, supported by the quantitative assessment summarised in the payback profiles.

3.3.1 Payback profiles – scoring consistency

One of our key concerns with the scoring system was the level of subjectivity in the valuations and hence the level of agreement that there would be between scorers. We felt that a high level of agreement between individual scorers would indicate that the system was reasonably robust and illustrate that different members of the team came to similar conclusions in valuing the outputs and outcomes from the case studies. In Table 3.4 we present the degree of disagreement in the second round scoring by payback category.

Initially, we used the original RAND/UCLA Appropriateness Method (RAM) algorithm to determine cases of disagreement. Using this definition of disagreement we never saw a case of disagreement in the scoring, even in the first round of scoring prior to the group discussion. Two reasons for this are that: (1) the RAM algorithm is designed to detect disagreements where scores are split between the extremes of the scale; and (2) it uses a fairly narrow definition of disagreement. In order to examine the level of disagreement in more detail we looked at two alternative measures of disagreement: we counted the number of scores that deviated from the median score by either two or more or three or more. Even using our widest definition of disagreement (deviation from the median by two or more) less than 10% of the scores were in disagreement. In each category we have 144 scores, representing nine team members each scoring 16 case studies.

Examining both our deviation-based measures of disagreement, we found that disagreement was highest in the first three payback categories, and had a similar value in all three of these categories. This is likely to reflect the subjectivity in valuing the variety of outputs and outcomes in these categories. The higher level of agreement in the final two categories probably reflects the high number of case studies where there was no output,

Table 3.4: Degree of disagreement in scoring case studies

	Overall disagreement	Knowledge production	Research targeting and capacity building	Informing policy and product development	Health and health sector benefits	Wider economic benefit
Number of scores	720	144	144	144	144	144
RAM	No disagreement	No disagreement	No disagreement	No disagreement	No disagreement	No disagreement
% of scores 3 or more from median	0.7	1.4	0.7	0.7	0	0
% of scores 2 or more from median	9.6	13	13	14	4.9	3.5

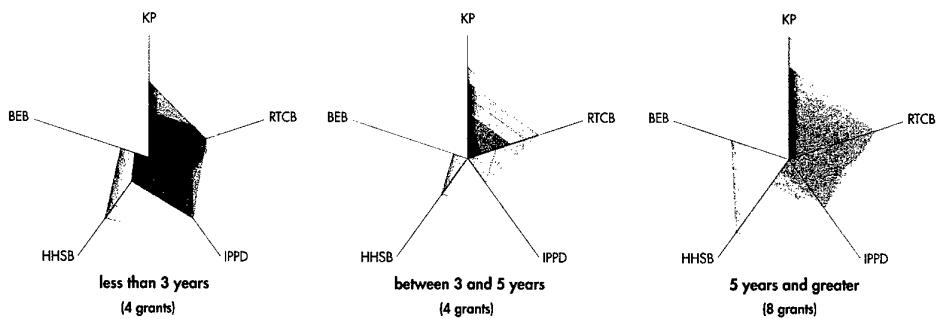
and hence a score of 1. Overall, we were surprised by the low level of disagreement between scorers. This suggests that such a scoring system does provide a useful method of summarising case study outputs to allow rapid comparisons of groups of case studies.

3.3.2 Possible confounding variables

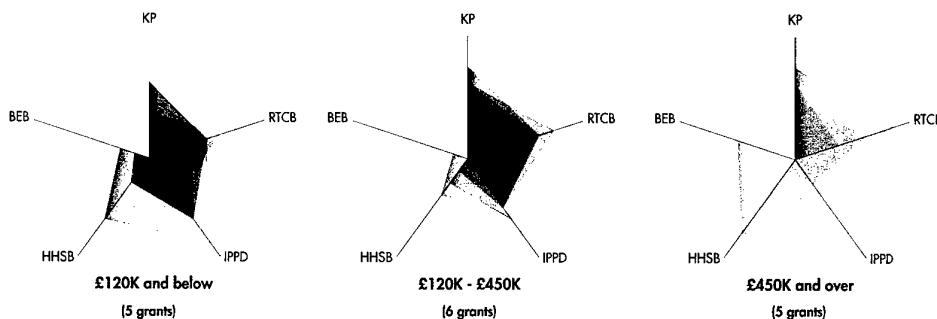
Given the small number of case studies, in making comparative assessments there is a risk that the analysis can be confounded by the inputs into the research, including the time and value of the grant and the experience of the researcher. Hence in Figure 3.2, we have summarised the payback from the grants by length of grant, value of grant, value per year of grant and postdoctoral experience. This information is most useful to put into context the other comparisons that follow; however, it is interesting to note that the shorter grants seem to provide a significant level of payback and that there is no obvious relationship between length of grant and level of payback. This may be partly because the scoring scale considered the payback from the grants in the context of the maximum payback that was likely from the proposed work.

Looking at absolute size of grant, again, there seems to be no obvious trend level of payback. Initially it appears as though the largest grants have the highest payback, but this is mainly due to case study L, which has by far the largest payback of any case study, so we conclude that the variability of payback seems to increase with grant size. A similar pattern is seen when comparing grants by annual cost, with the larger grants providing a wider range of amounts of payback.

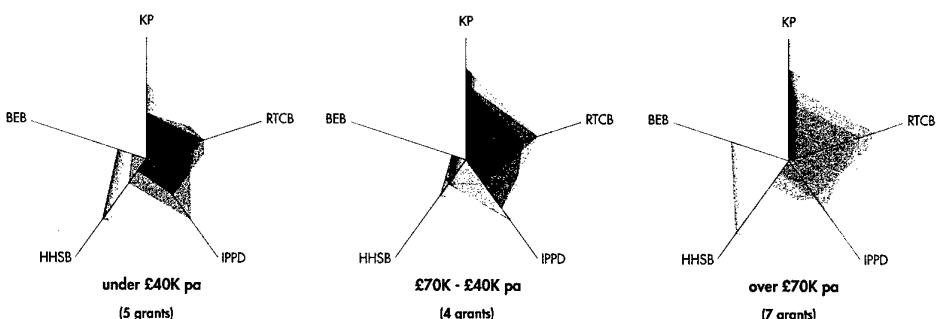
Another independent variable that we thought might affect the translation was the experience level of the researcher. In order to try and quantify this we used a measure of "research age", which was the number of years between a PI gaining their first research degree and the start of their case study grant. Research age varied widely within the case studies: the lowest ages were in fact negative as three of the researchers gained their first research degree after the start of their grants, and the highest research age in the sample was around 27. Given the differences in career trajectory between basic, clinical and AHP scientists this measure may not be an ideal way of quantifying research experience; however, it gives us a useful approximation. It seems that the level of translation, especially into the later payback categories, increases slightly with increasing research age.



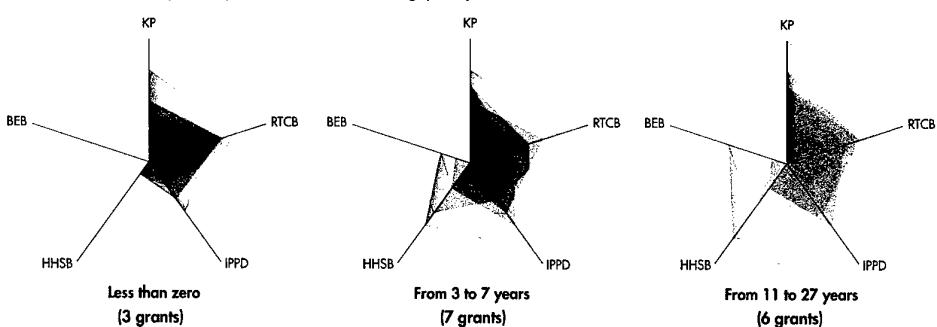
Case studies grouped by length of funding (institutes counted as five years or greater)



Case studies grouped by total amount of funding



Case studies grouped by amount of funding per year



Case studies grouped by research age of PI

Figure 3.2: Payback profiles grouped by potential confounding factors

3.3.3 Mode of funding

Four modes of funding provided by arc (in the early 1990s) were considered in the context of this study: project grants, programme grants, fellowship grants and institute grants. As summarised in Box 2.2, project grants are designed for work on a specific research question over the duration of around three years; programme grants support a portfolio of more speculative long-term research over a period of five years; fellowships serve to develop the career of individual; and institute grants help to maintain a centre of RA research with both clinical and basic strands, for four to five years at a time.

Analysis of Table 3.5 and Figure 3.3 suggests that, despite lower costs and shorter duration, at a median value of £90,000 over 2.5 years, project grants yield considerable payback over a range of categories, and do not appear to underperform the other three funding modes (programme grants at a median value of approximately £480,000 over five years, fellowships at approximately £50,000 over five years, institute grants at approximately £450,000 over four years). Considering their lower level of funding this is a considerable achievement.

Funding provided to institutes resulted in the widest range of levels of payback and also encompassed the highest scoring study. This suggests that institute funding allows the pursuit of speculative and innovative research that, when successful, has large payback. Project grants and institute funding include the case studies with the broadest health benefits (case studies K and L respectively).

By comparison, programme grants (roughly comparable to institute grants in median funding amounts and periods) perform poorly, with the payback recorded as limited to two PhDs, a guideline citation and a clinical test. However, it should be borne in mind that we only examined three programme grants and that the award system for such grants has become far more rigorous and competitive since the early 1990s when these grants were awarded.

Table 3.5: Qualitative analysis of payback by mode of funding

	Project	Programme	Fellowship	Institute
Knowledge production	56 papers with a total of 187 citations per year	82 papers with a total of 126 citations per year	63 papers with a total of 196 citations per year	101 papers with a total of 466 citations per year
Research targeting and capacity building	4 PhD/MDS awarded from work on grant [case studies A, C, K and J] Provision of reagents [case study C] Japanese study on occupational lifting [case study B] Informed subsequent work looking at the effect of APL antibodies on in vitro fertilisation [case study K] Informed >£2 million MRC RCT [case study A] Development of mouse model [case study D]	2 PhD/MDS awarded from work on grant [case study F]	9 PhD/MDS awarded from work on grants [case studies G, H and I] Contributed to the establishment of MRC for Immune Regulation [case study H]	13 PhD/MDS awarded from work on grants [case studies I, O and P] Development of technological know-how in genetic mapping [case study O] Informed shift in the use of biologicals as therapeutic targets [case study I]
Informing policy and product development	Citation on systematic reviews [case studies A and B] Directly supported RCOG guideline recommending a combination therapy of aspirin and heparin for women with antiphospholipid syndrome [APS] [case study K] Cited on Dutch guideline [case study B] Cited on IIAC assessment of whether hip OA informers should be a prescribed disease [case study B] Informed management of ACLD [case study J] Combined treatment of aspirin and heparin reduced the risk of miscarriage for women with APS [case study K] Study results improved the management of back pain patients through study days run by PI [case study A] Improved rehabilitation of ACLD patients [case study J]	Cited in European guideline [case study E] Test for chondrocytasia type Schmid [case study F]	Test for SLE [case study G] Informed development of a class of drugs to treat DVT [case study H]	Widely cited in a number of guidelines in setting context [case study P] Tested hypothesis that RA could be treated using TNF alpha inhibitors [case study I]
Health and health sector benefits	Unquantified return in reduction of days of work [case studies A, B, J and K]			70% patients treated with anti-TNF leads to significant health improvement [case study I] Understanding of prevalence of arthritic disease [case study P]
Wider economic benefits				Economic return to companies licensed to produce drug [case study I]
Confounding factors	Median value of grant Median length of grant Median research age	c. £480,000 2.5 years 3.5 years	c£250,000 5 years 7 years	c. £450,000 4 years 16.5 years

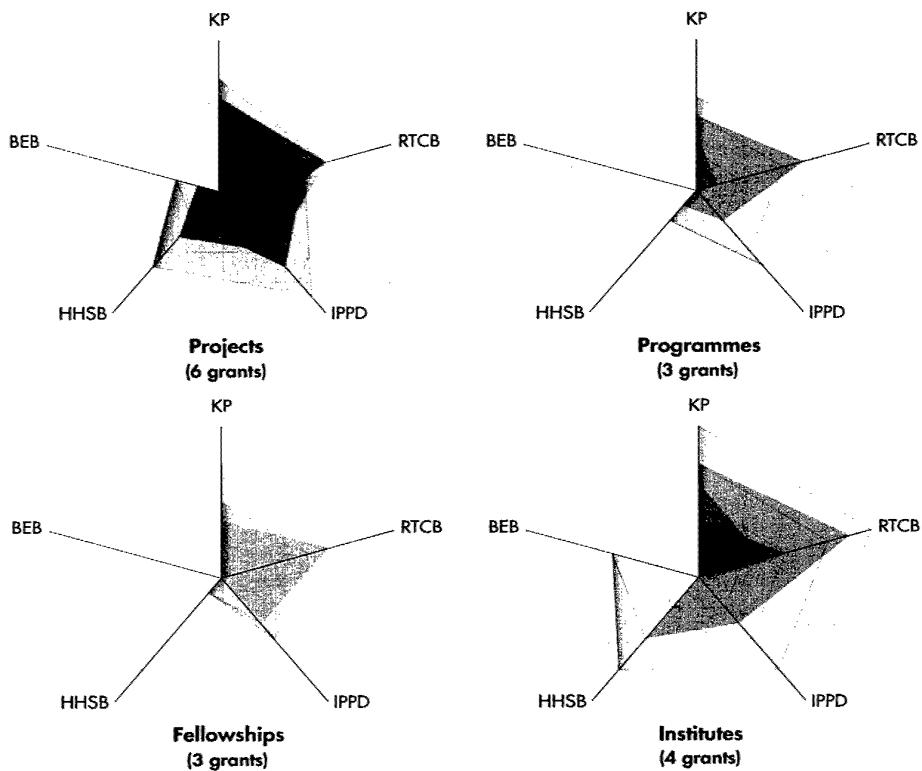


Figure 3.3: Payback profiles by mode of funding

3.3.4 Type of research

arc research funding can be divided into three broad areas of research: basic research, clinical research and AHP research. We classified the research grants according to the qualifications of the PI (as explained in Appendix A). This allows us to compare the payback profiles from research in the three different areas (Figure 3.4) and also to consider the qualitative comparison (Table 3.6).

It is notable that for all forms of research, translation occurs, and there are examples of the later categories of payback. Basic research appears to be more consistent in producing high levels of knowledge, which feed into research targeting and capacity building.

As might be expected, the payback from basic research declines more rapidly than clinical research as the course of translation is followed towards health benefits. It is not clear how much of this difference in the extent of payback is due to the need for a longer timeframe for the translation of basic research, which would move it past the timescale of our study. For example, the identification of TNF-alpha as an important signalling molecule in cartilage damage was reported from basic research in 1986 (Saklatvala 1986); however, we followed this story from 1992 as a clinical research project. Translation of basic research also feeds primarily into drug development or diagnostic tests, whereas clinical research produces a wider range of outputs including citation in systematic reviews and clinical guidelines.

Possibly due to these timescale issues, the clinical research studies appear to have the largest amount of payback. Considering the profiles, the difference between the outputs of

basic and clinical case studies is lower if the development of anti-TNF is excluded from consideration; however, from a qualitative point of view the outputs of clinical case studies still include a wider variety of clinical outputs.

From the two case studies examined, AHP research appears to be translated more effectively – with less decrease in payback when moving around the payback profile. AHP research also includes one of the few examples where the influence on policy was scored as higher than the level of knowledge production.

It should also be borne in mind that amounts and periods of funding varied across the three areas, with clinical projects emerging as the most costly of options (median value of £240,000 over 2.5 years), compared to basic science at £145,000 over four years and AHP at £100,000 over 2.5 years. Again, this underlines the effectiveness of translation in the AHP area.

3.3.5 Bibliometric impact of researcher

To investigate whether bibliometric impact was related to level of translation we compared our “High” and “Mid” impact researchers – with rankings determined independently for the basic, clinical and AHP researchers. The researchers had been allocated to these categories based on a bibliometrics assessment of their publication record from the start of the selection window (1990) up to almost the present day (2002).

Again, it was clear that translation was occurring both by “High” and “Mid” impact researchers. Although “Mid” impact researchers tended to produce less knowledge, they appeared to translate this knowledge more effectively than the “High” impact researchers: notice how the profiles of the “Mid” researchers “catches up” the “High” impact researchers on the IPPD, HHSB and BEB axes. Furthermore, it should be borne in mind that “High” impact researchers received, on average, considerably more funding over a longer period (median £400,000 over five years) than “Mid” impact researchers (median £105,000 over three years). An additional confounding factor may be the tendency of our impact classification to reflect a split between senior and junior researchers rather than overall career success.

Table 3.6: Qualitative analysis of payback by type of research

	Clinical	APHP
Basic		
Knowledge production	100 papers with a total of 372 citations per year. 20 PhD/MIDs awarded from work on grant (case studies C, F, H, O and P) Provision of reagents (case study C) Contributed to establishment of MRC for Immune Regulation (case study H) Development of technological know-how in genetic mapping (case study C) Development of mouse model (case study D)	1.83 papers with a total of 588 citations per year. 6 PhD/MIDs awarded on grant (case studies G, I and K) Informed Japanese study on occupational lifting (case study B) Informed subsequent work looking at the effect of APL antibodies on <i>in vitro</i> fertilisation (case study K) Informed shift in the use of biologicals as therapeutic targets (case study L)
Informing policy and product development	Informed development of a class of drugs to treat DVT (case study H) Test for chondrocytasia type Schmid (case study F)	Citation on systematic reviews (case study A) Directly supported RCOG guideline recommending a combination therapy of aspirin and heparin for women with APS (case study K) Cited in Dutch guideline (case study B) Cited in European guideline (case study E) Cited in IAC assessment of whether hip OA in farmers should be a prescribed disease (case study B) Widely cited on a number of guidelines (case study P) Test for SLE (case study G) Tested hypothesis that RA could be treated using TNF alpha inhibitors (case study L)
Health and health sector benefits		Study results improved the management of back pain patients through study days run by PI (case study A) Improved rehabilitation of ACD patients (case study J)
Wider economic benefits		Unquantified return in reduction of days off work (case studies A and B) Economic return to companies licensed to produce drug (case study I)
Possibly confounding factors	Median value of grant Median length of grant Median "research age"	c. £240,000 c. 4 years 11 years
		c. £100,000 2.5 years -2 years

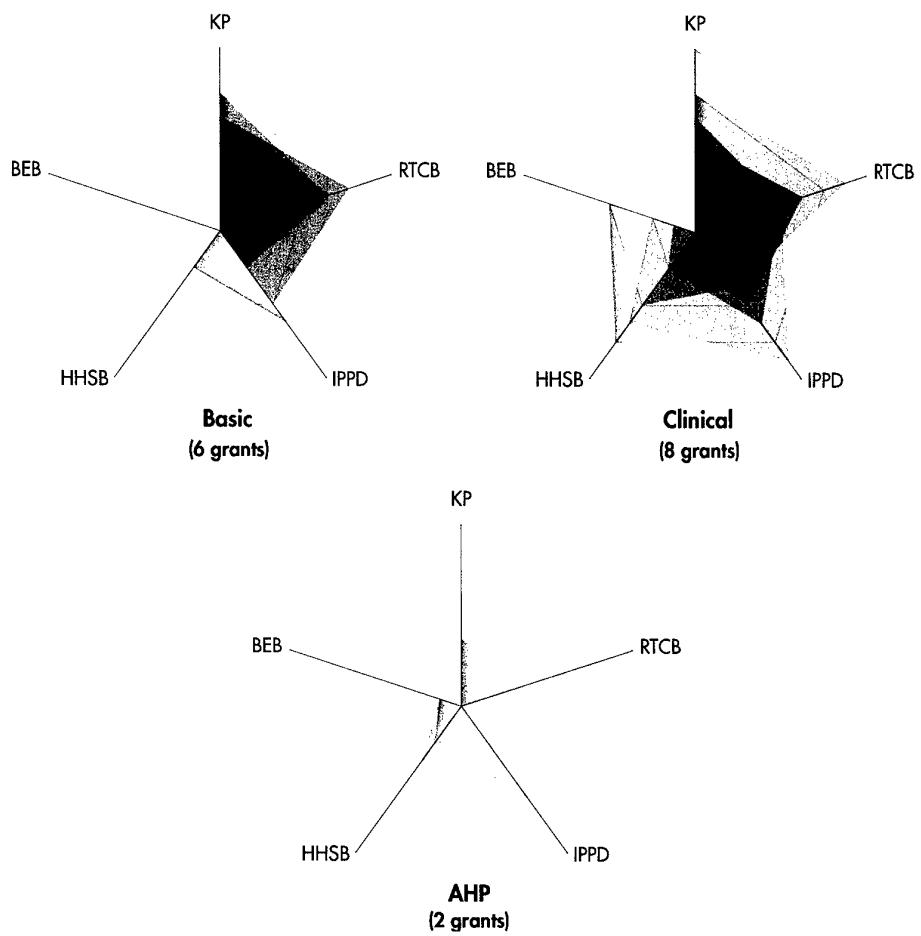


Figure 3.4: Payback profiles by area of funding

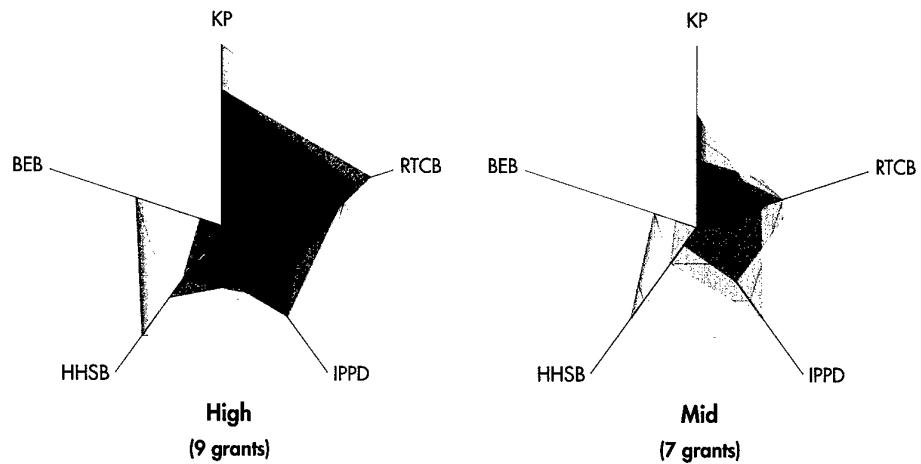


Figure 3.5: Payback profiles by bibliometric impact of researcher

3.3.6 Knowledge production as a predictor of health and health sector benefit

Our final cross-case comparison was to examine whether the level of knowledge production could be used as a predictor of the level of translation, and consequently the health and health sector benefit payback from a grant. To do this we grouped the case studies by their score for knowledge production and examined each group's level of health and health sector benefit payback.

While it is clear that there is some tendency for case studies that produce large amounts of knowledge to translate effectively in order to produce health benefit, there are also case studies that produce large amounts of knowledge but little payback in terms of health benefit. To investigate the relationship further, we treated the scores as continuous variables and calculated the correlation between knowledge production and health benefit. The correlation coefficient was $R^2 = 0.33$, which confirms the visual suggestion of a moderate correlation. This implies that there are factors other than simply the level of knowledge production that determine the level of translation and hence the final payback in terms of health and health sector benefit.

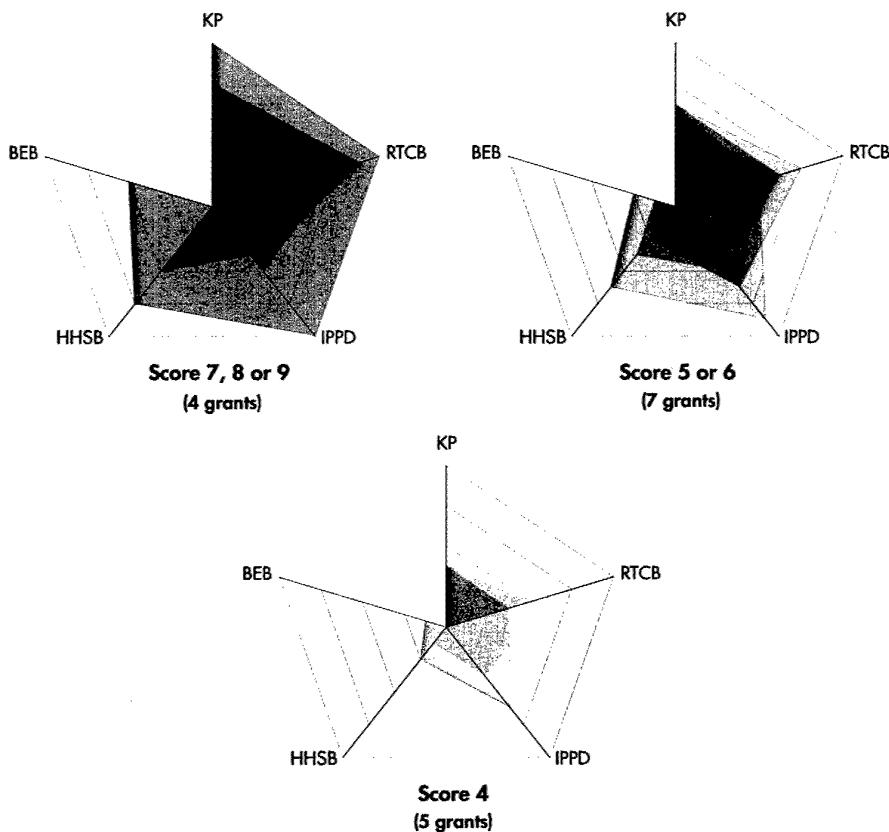


Figure 3.6: Payback profiles by level of knowledge production

3.4 Closing remarks

In this chapter we have presented the results of our evaluation considering the variety and scale of payback by model stage and overall payback by category. We then used this information to carry out comparisons between different modes and areas of funding. In the next, and final, chapter we consider the policy implications of our findings.

Table 3.7: Qualitative analysis of payback by track record of researcher

	High impact	Mid-impact
Knowledge production	212 papers with a total of 769 citations per year	90 papers with a total of 206 citations per year
Research targeting and capacity building	21 PhD/MDs awarded from work on grant [case studies A, C, F, G, H, I and P] Provision of reagents [case study C] Informed >£2 million MRC randomised controlled trial [case study A] Contributed to the establishment of MRC for Immune Regulation [case study H] Informed Japanese study on occupational lifting [case study B] Informed shift in the use of biologicals as therapeutic targets [case study I]	7 PhD/MDs awarded on grant [case studies I, J and K] Informed subsequent work looking at the effect of APL antibodies on <i>in vitro</i> fertilisation [case study K] Development of technological know-how in genetic mapping [case study O] Development of mouse model [case study D]
Informing policy and product development	Citation on systematic reviews [case studies A and B] Cited in Dutch guideline [case study B] Cited in European guideline [case study E] Cited in IIAC assessment of whether hip OA in farmers should be a prescribed disease [case study B] Informed development of a class of drugs to treat DVT [case study H] Test for chondrolysis type Schmid [case study F] Test for SLE [case study G] Tested hypothesis that RA could be treated using TNF alpha inhibitors [case study I] Widely cited in a number of guidelines [case study P]	Informed management of ACLD [case study J] Directly supported RCOG guideline recommending a combination therapy of aspirin and heparin for women with APS [case study K]
Health and health sector benefits	Study results improved the management of back pain patients through study days run by PI [case study A] 70% patients treated with anti-TNF leads to significant health improvement [case study I] Understanding of prevalence of arthritic disease [case study P]	Combined treatment of aspirin and heparin reduced the risk of miscarriage for women with APS [case study K] Improved rehabilitation of ACLD patients [case study J]
Wider economic benefits	Unquantified return in reduction of days off work [case studies A and B] Economic return to companies licensed to produce drug [case study I]	Unquantified return in reduction of days off work [case studies J and K]
Possibly confounding factors		
Median value of grant	c. £400,000	c. £105,000
Median length of grant	5 years	3 years
Median research age	7 years	3 years

The aim of this study was to develop a system for evaluating the long-term outcomes of arthritis research, with a view to helping the Arthritis Research Campaign (arc) manage and encourage the successful translation of research outputs into outcomes that benefit people with arthritis. In this chapter we provide a brief overview of our work and draw out a number of conclusions, implications and policy recommendations for arc.

4.1 Overview

To structure our analysis we adopted and refined the Buxton and Hanney payback framework that conceptualises the relationship between research inputs, processes, outputs and outcomes. As discussed in Chapter 1, the use of logic models is one element of the research evaluator's "toolkit" (Table 1.1); this provided a consistent structure to develop 16 case studies and allow cross-case analysis.

In Chapter 2 we explained our approach in detail. This included a detailed description of the payback framework, including an examination of the five payback categories and the seven-stage, two-interface, payback model. In addition, we discussed the process by which we selected our 16 case studies from a possible 556 grants, along with the data collection methods that we employed. Finally, we explained how we undertook the cross-case analysis. This included a novel way of scoring the payback from case studies and tests for levels of agreement.

In Chapter 3 we presented our results. This included a comprehensive cataloguing of the payback for each of the 16 case studies by the five categories and an analysis of the payback model. In addition, we made a series of qualitative and quantitative cross-case study comparisons, assessing, for example, the payback from project grants versus programme grants.

On the basis of the evidence presented in Chapter 3, in this chapter we present our findings as a series of six policy observations. For each observation we explain the issue, the supporting evidence and the strength of that evidence from the case studies. Where appropriate we assess the implications of the observation in light of the existing science policy literature. We conclude by making a recommendation to arc regarding how they could begin to address the identified issue. However, before presenting our observations, we identify and discuss a number of caveats regarding our approach and analysis.

4.2 Study limitations and caveats

This study has shown how the returns from arthritis research, or any other field of research, can be evaluated. However, there are several limitations to our approach. By highlighting these issues we do not wish to undermine the importance of the findings that we present in this report, but to illustrate some of the challenges faced in evaluating research.

4.2.1 Linearity of the payback model

It is acknowledged that the payback model oversimplifies the way in which research is conducted and it is important to bear in mind that such models necessarily abstract from various feedback loops and secondary effects. Furthermore, the process by which science translates “from bench to bedside” is not as linear as depicted in the model. Nevertheless, the advantage of the payback model (and, for that matter, all logic models) is that it provides a workable framework within which to evaluate the outputs and outcomes of research.

4.2.2 Use and generalisation of case studies

A second issue is the use of, and generalisation from, case studies. Given the resource-intensive, in-depth approach that we adopted, the number of case studies had to be restricted. This immediately raises issues about whether the study can be generalised and becomes a debate about the legitimacy of qualitative research and the use of case studies in general. Yin, who has written the seminal text on case study research, argues that “the case study is but one of several ways of doing social science research ... [each with] peculiar advantages and disadvantages” (Yin 2003, p. 1). He goes on to state that the case study is the preferred method when explanatory questions, such as how research is translated, are to be addressed; Yin himself used a case study approach when examining research utilisation (Yin and Moore 1988).

In terms of scientific generalisation, Yin makes the point that “scientific facts are rarely based on single experiments” and that the “investigator’s goal is to expand and generalise theories (analytic generalisation) and not to enumerate frequencies (statistical generalisation)” (Yin 2003, p. 10). By using a multiple-case study approach (as adopted here) the evidence is made more compelling; the analogy being with experimental replication, rather than statistical sampling.

4.2.3 Selection of case studies

Related to the issues of generalisation from case studies are the issues of biases in the selection of our grants. With a limited number of case studies we were keen to ensure that those we investigated were tractable and interesting in terms of the translation of research. We also needed them to mirror the diversity of arc funding. To achieve this, we used a selection matrix to ensure variety coupled with a final subjective selection step by the arc Development Committee. The most obvious selection bias in our sample is our examination of “High” and “Mid” impact investigators and our exclusion of “Low” impact investigators. Because of this selection procedure we have been careful to consider the validity of generalisation from our findings and this has affected the wording of our conclusions and recommendations, for example, in our consideration of modes of funding.

4.2.4 Attribution and analysis of payback

The attribution of payback to specific research grants or funders is an issue central to research evaluation. This issue was explored in some depth at a 1999 international workshop on research evaluation (Buxton, Croxson and others 1999; Croxson and others

2001), and it was concluded that there are many complications in identifying the impact of specific funding. How far it is desirable to attempt this in detail depends partly on the purposes for which the evaluation is being conducted (Buxton, Croxson and others 1999; Croxson and others 2001). It is undoubtedly true that there are a range of unquantified and (largely) unspecified non-arc inputs into the research evaluated in the 16 case studies. However, a crucial point we would highlight is to consider of whether any of the benefits identified would have arisen without the specific arc funding.

In terms of analysis, the major caveat regarding the quantitative aspect of the approach reported here is the robustness and consistency of the novel scoring system that we developed for the payback profiles. Although we believe that this system is a major step towards being able to operationalise and embed a multidimensional research monitoring and evaluation system, we would stress that it is developmental. As noted in Chapter 2 (section 2.5.2 and discussed below), we need to have a better understanding of the reliability of the scoring system, a tighter definition of the scoring scale and an understanding of the quantity and type of information necessary to produce informed scores.

4.2.5 Elapsed timescale

In deciding on a time window for the evaluation, a compromise needed to be found between the quality of records, the ability of researchers to recall their activities and allowing enough time for research outputs to develop (Bozeman and Kingsley 1997). In this study we have used a 10-year lag; this did not seem to create many difficulties in recall by the researchers and there were plentiful paper records available covering this timeframe. The 10-year lag also seems to have allowed us to identify a sufficient variety of research outputs and outcomes to draw some conclusions with respect to the strengths and weaknesses of different modes and types of funding. However, we consider that it was not long enough to allow a fair comparison between basic and clinical research. Figure 3.4 could be taken to suggest the superior paybacks of clinical research, however there are a number of potential – but unrealised – paybacks from a number of the case study grants that we evaluated, particularly the more basic ones. Previous work examining how long it takes for basic research to inform clinical guidelines suggests that even after four generations of citation, taking around 17 years less than 20% of research is basic (Grant and others 2000), and it is worth noting that the anti-TNF work would have been basic research in the mid-1980s.

4.3 Policy observations and recommendations

Based on our analysis of 16 case studies we have drawn out six policy observations and recommendations for arc. We believe that these findings will be of interest to other research funders and evaluators who are concerned with measuring the impact of science.

4.3.1 There is a diversity of research payback

There is strong evidence from our analysis that there is a considerable range of research paybacks and that many of these would not have been identified without the structured, in-depth case study approach that was employed in this evaluation.

In Table 4.1 we highlight the various types of payback identified from the 16 case studies. While this is undoubtedly an impressive list, it should be remembered that our sample of 16 case studies were selected purposively and biased towards more successful investigators. Nevertheless, it is worth emphasising that these returns come from only 16

of 556 (or 3%) possible grants, and even if cautiously extrapolated, illustrate a diverse and significant return.

Given this diversity of outcome it is important that **arc** uses an assessment method that captures a broad range of payback. The payback categories – or iteration thereof – developed in this study would provide a structured basis for any further monitoring or evaluation of the outputs and outcomes of research funded by **arc**. This is important, as there is a tendency in research evaluation to rely solely on bibliometric analysis and, as shown in this study, this can seriously underestimate payback.

We have shown also that it is important to look over a long timescale to pick up this variety of paybacks. In order to track and catalogue such paybacks routinely it would be important to use an incremental database that could be added to over time, and to return to grants periodically after they had finished to update their paybacks.

The importance of a multifaceted approach to research evaluation was confirmed by Cronin and Normand (1998) in their review of the literature. First, they concluded that there was consensus that all criteria needed a qualitative and quantitative information base; second, that a number of different criteria should be used. This was best articulated by Martin and Irvine (1983) who proposed the concept of “converging partial indicators”, that is, a series of incomplete indicators that indicate that the reliability of the payback assessment when they are all pointing in the same direction.

Based on the evidence of this study, and on that reported in the literature, it is recommended that **arc** *should survey all forms of payback when monitoring and evaluating, and should look to build up a long-term picture of the returns from arthritis research.*

4.3.2 Individuals translate research

There is good evidence from our 16 case studies that when translation occurs, in both non-commercial and commercial settings, it is largely due to the conviction, effort and personal networks of the PI. Our cross-case comparisons lead us to the conclusion that translation is not associated with the type of research, mode of funding or the bibliometric impact of the PI. Examining the case study narratives leads us to the conclusion that it is the PI that is the key factor in translation.

In addition to analysing the paybacks that occurred, we also analysed the pathways by which this translation took place, these are shown in Figure 4.1. In this figure we have superimposed them onto the payback stages in which they occur. We identified a variety of knowledge flows, shown with arrows, and two main routes of translation for the case studies (highlighted in red).

Table 4.1: Summary of research paybacks

Payback category	Payback	Example
Knowledge production	Peer-reviewed publications in the serial literature	302 papers receiving a total of 975 citations per year attributable to case studies
Research targeting and research capacity	Postgraduate research training Subsequent career development of principal investigators (PIs) and research assistants The transfer of technical know-how Informing future research studies	28 PhD/MDs from work on the case studies Development of technological know-how in genetic mapping Informed >£2 million Medical Research Council (MRC) randomised controlled trial Use of biologicals as therapeutic targets
Informing policy and product development	Informing recommendations in clinical guidelines and other policy advice Informed development of clinical tests	Recommendation in Royal College of Obstetricians and Gynaecologists guideline on the use of aspirin and heparin for women with antiphospholipid syndrome (APS) Recommendation in IIAC assessment for hip osteoarthritis (hip OA) in farmers to be a prescribed disease Clinical test for a rare type of systemic lupus erythematosus (SLE) and chondrodysplasia type Schmidt
Health and health sector benefits	Improving the quality of life for people with rheumatoid arthritis (RA) Reducing the likelihood of recurrent miscarriages for women with APS	Hundreds of thousands of patients treated with anti-TNF of whom 70% experience a significant improvement in health Use of aspirin and heparin for women with APS increases live birth rate by 40% compared to the use of aspirin alone and by 60% compared to no treatment at all.
Wider economic benefits	Unquantified economic returns resulting from a reduction in days off work and sales of licensed drugs	

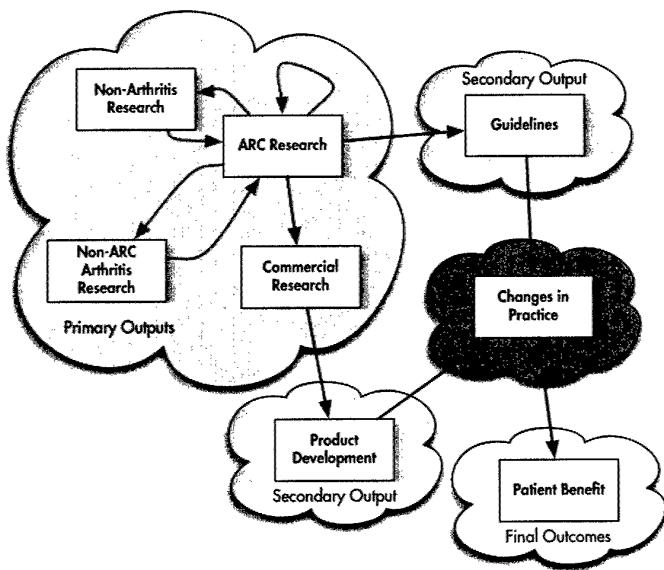


Figure 4.1: Payback pathways for arc-funded research

The first route, used by both clinical and basic research, was via industrial development with the case study research informing commercial research programmes (case studies A, H and L); this could then lead to product development, adoption and health benefit. In this pathway the movement of the research from an academic environment into the commercial world was an important hurdle which was not surmounted by case studies D and E. The key factor in overcoming the hurdle appeared to be the networks and contacts of the PI – in all cases of successful translation the PI had a close and friendly relationship with an industrial collaborator.

The second route of translation, followed by clinical and AHP research (case studies A, B, E, J and K), was by incorporation into guidelines and systematic reviews. Again, this route seemed to depend heavily on personal networks – where two case studies were cited in clinical guidelines, the writing of the clinical guidelines had involved the PI.

In addition to the two major routes found in the case studies, and described above, there are likely to be other routes that we did not see in our sample of grants, for example, from the laboratory to the clinic as a phase I clinical trial. One example of another route found in the case studies was the development and implementation of a diagnostic test based on the work in case study F, which was initiated and carried out by a European Union (EU)-funded genetic testing network. This illustrates that there are further pathways of translation that should be looked for in future work.

One of the key points emerging from the literature is that partnership between researchers, practitioners, policymakers and industrialists is often a precondition to successful translation, particularly in relation to research that is conducted for government departments (NAO 2003a). Underpinning this is the need for good communication between researchers and research users, and the importance of developing long-term relationships (Hanney, Gonzalez-Block and others 2003; Innvar and others 2002). However, such partnerships often do not occur spontaneously. An approach that has been

applied successfully to health service research is the model of "Linkage and Exchange" used by the Canadian Health Services Research Foundation (Lomas 2000). The basis of this model is that bringing policymakers, who can use the results of a particular piece of research, into its formulation and conduct is the best predictor for seeing its successful application.¹⁸ It is suggested that this concept of collaborative research results from convergent evolution through both the academic and policy worlds (Denis and Lomas 2003).

As *arc*'s diverse research portfolio extends beyond health services research, the explicit Linkage and Exchange Model may not be directly applicable, especially for the more basic biomedical research. Nevertheless, attention is focusing increasingly on the role of researchers in translating findings, irrespective of any involvement of policymakers in the commissioning of research (Ferlie and Wood 2003). The importance of networks is noted across the spectrum, from health-oriented social science research (Molas-Gallart and others 2000) to research that is relevant to the pharmaceutical industry (Albertini and Butler 1995).

On the basis of the evidence arising from this study and from others, which has been discussed above, we recommend that *arc* *should selectively support investigators in translating their research*.

We suspect that not all investigators will have the skills to drive the translation of their work; due to this we see the primary purpose of this support not in providing incentives for translation, but in supporting those with the necessary skills and recognising their activities. A secondary benefit may be to interest additional PIs in the process of translation.

Considering the experience from the case studies we suggest that this support could take two forms:

translation awards to promote the successful transfer of knowledge with potential health benefit;

interaction awards to develop productive relationships between researchers and policymakers or industry.

Translation awards would be project-focused and awarded selectively to PIs who had generated promising results from their *arc*-funded research. The grants would support activities designed to disseminate and translate the research findings, for example, resources for the dissemination of reagents, running study days for a proven intervention, or developing a business case for the commercialisation of a therapeutic target. The award criteria would focus on:

- evidence of the PIs' skills in translation;
- the potential return or payback from translation;
- the stated route or plan of translation; and
- relevance to *arc*'s strategic aims.

The grants would *not* be awarded explicitly on the basis of scientific quality (as this would have been reviewed in the award of the original research grant).

¹⁸ See <http://www.chsrf.ca> for more detail.

Interaction awards would be people-focused and provide resources to help **arc**-funded researchers develop networks with potential users of their research. This could include supporting secondments to or from commercial and non-commercial laboratories, and participation in policymaking networks. The award criteria would be similar to those for translation awards but also would examine evidence of the PIs' existing networks.

To stimulate translation we suggest that *awards be made in both reactive and directed modes*. In reactive mode, PIs will be asked to apply for the awards in a similar way to current scientific grants. In the directed mode **arc** (possibly through the Development Committee) would actively nominate investigators for awards, based on assessment of their grant reports and publications.

4.3.3 Short focused project grants seem to provide value for money

There is good evidence from our qualitative and quantitative analysis that project grants provide value for money when compared to programme grants, fellowships and institutes. The payback arising from project grants is surprisingly similar to that arising from the other modes of funding, while the median value of a project grant is £90,000 compared to £250,000 for fellowships, £480,000 for programmes and £450,000 for institutes.

The cross-case analysis presented in Figure 3.3 and Table 3.5 shows a significant and diverse range of payback arising from the six project grants that we examined. This included, for example, 62 papers, four PhD/MDs, informing four clinical guidelines or policy documents, reduced risks in miscarriages for women with APS, and unquantified reductions in days of work. By comparison, the three programme grants resulted in 82 papers, two PhDs, one citation in a clinical guideline and the development of a clinical test for chondrodysplasia type Schmidt. This does not appear to show the much higher returns that might have been expected from programme grants.

Of all the observations we have made from our analysis, this was the most unexpected and surprising. There is a widespread view in science policy that long-term stable funding is preferential, although Narin's (1989) study showed that all types of support mechanisms contributed significantly to 13 clinical advances in cancer research. However, there are a number of potential confounding factors that mean that this conclusion needs to be treated with caution. For example, four of the six project grants were patient-focused, and three of the project grants involved a randomised controlled trial. This suggests perhaps that the projects were more "mission-oriented" and therefore likely from the outset to lead to faster payback. Conversely, the fact that patient-focused, randomised controlled trial-type work is being funded through projects illustrates the importance of maintaining a mechanism for funding short-term focused research of this nature. This leads us to recommend that **arc** should continue to support project grants as part of its wider funding portfolio.

4.3.4 Intended and unintended flexibility is used advantageously

There is some evidence from our case studies that investigators successfully exploit flexibility in the scientific and administrative management of grants.

Flexibility was observed in a number of different guises. For example, as is normal research practice, investigators were given a free hand in the recruitment of research assistants. What was interesting, however, was how this freedom was used successfully to recruit individuals with unusual backgrounds, for example, in human resource management to run a clinical trial (case study B). It was also evident from reviewing the correspondence between **arc** and the PIs that a number of personnel changes were made on grants and again, the PIs had considerable flexibility in how this was managed and

implemented. Finally, in a number of cases, the research objectives changed either because the proposed work occurred quicker than anticipated (case study F) or because the work or strands of the work were unachievable (case studies B and D). In both scenarios, the PIs productively pursued related but different lines of research.

In none of the case studies was there any evidence that this flexibility had a negative effect on the scientific outputs and outcomes of the research. Therefore, this observation supports the continuation of **arc**'s current policy of flexibility in funding.

One radical approach which **arc** may wish to consider is the introduction of fixed budget grants of, for example, £50,000, £100,000, £250,000 and £500,000. Under such a system, investigators would not be required to produce "zero-based" budgets, as is current practice, but would be asked to illustrate how they would plan to spend a fixed sum in order to meet their proposed scientific objectives. The advantage of such an approach is that it gives the investigator considerable flexibility in undertaking the research. In addition, there is potential to reduce the transaction costs associated with the management and administration of a grant. The disadvantage to this is that **arc** would surrender some control over the way in which its research funds are spent (although this could be mitigated by the introduction of appropriate guidance and audit) and that university finance departments may be unable or unwilling to operate such a system.

Given the above, we recommend that **arc** *should maintain its flexible approach to the funding and administration of research grants, including reviewing the costs and benefits of fixed budget grants.*

4.3.5 Referees' contributions to the peer-review process are of variable benefit

There is some evidence from our case studies that for successful applications, referees' contributions to review panels do not add significant scientific value to reviewed proposals.

In only five out of the 16 case studies did the PIs recall the reviewers' comments improving the project. These improvements were seen as being of greatest value where there was a dialogue between the referees and the investigators, particularly in the two cases of resubmission where the investigator was asked to reapply after addressing the referees' concerns (case studies A and I). In two cases (M and O) the grant was reduced, with extension conditional upon progress at a given point in the future. For the final two cases (K and L), if the referees' comments had been taken at face value and had not been overruled by the assessing panel, then the proposed work would not have been funded. It is worth noting that the referees' concern was due to whether the proposals were within **arc**'s remit, rather than regarding their scientific quality. Nevertheless, these two cases had the highest impact in terms of health and health sector benefits.

These observations needs to be treated with care for number of reasons. First, as a result of the study design, we looked primarily at those grants that had been awarded. Second, **arc** only recorded the outcome of funding decisions in panel minutes and there is no explanation as to why referees' comments were ignored or overruled. Finally, the peer-review system which operated in the early 1990s is very different from that which happens today.

The role of peer review is central to scientific decision-making but is not without criticism (Grant and Allen 1999). It has been suggested that the process is inherently conservative and is biased against speculative and innovative research (Wessely 1998). Horrobin (1990, 1996) has argued that innovation comes from unexpected observations which

often are opposed initially by experts, citing a number of supporting but anecdotal examples including the identification of B lymphocytes, and the introduction of *in vitro* fertilisation (IVF). From arc's perspective it is important to know whether innovative and speculative research is associated with translation (a topic that the current study has not addressed). Even if innovative and speculative research is not associated with translation, and given the lack of existing evidence surrounding the arc peer-review process, we recommend that arc *should review its peer-review processes to maximise their efficiency and effectiveness*, with the aim of improving their usefulness to arc and value to successful applicants.

4.3.6 The payback framework could be operationalised and embedded by arc

There is good evidence from this study that the payback framework developed for arc works and, given the appropriate management information, could be operationalised to prospectively stimulate and manage the returns from arthritis research.

The payback framework has proved to be effective in capturing the diverse range of research outputs and outcomes, and in identifying the cases where the research had been translated to benefit people with arthritis. If applied prospectively, the framework could be used to inform the granting of the recommended translation and partnership awards. Below we describe how arc could operationalise and embed the payback framework, and identify (in Box 4.1) a number of issues that need to be resolved prior to implementation.

We envisage a system that could have four different evaluation points. At year 0, the proposed grant would be peer-reviewed according to arc's current practice. That is, the PI will write an application, which will be commented on by referees. The assessing panel will review then both the application and referees' input and make a funding decision based on scientific quality. Running in parallel with this "quality review" we suggest that arc conducts a "translation review".¹⁹ This will involve the applicant and referees scoring the proposal on potential payback for each of the five categories on a scale of 1 to 9. The assessing panel would use this information, and their expert judgement, to rescore the proposal and the median values will be used to generate a "payback profile" for the application. For successful proposals, this information will be updated at regular intervals – we suggest at three, six and 10 years. Ideally, we would suggest that the PI, referees and panel do the updating, although in practice this may not be feasible. The updated portfolio of payback profiles will be monitored then by arc and used to inform the granting of translation and partnership awards. In time, this will allow arc to build a prospective database of research outputs and outcomes with which it can evaluate its overall impact and assess the effectiveness of its funding. We therefore conclude that arc *should consider developing systems for the ongoing and prospective monitoring and evaluation of its funded research*.

If arc decided to move to such a system, a number of outstanding issues would need to be resolved, as summarised in Box 4.1. The first issue is to have a better understanding of the reliability of the scoring system. It would be quite simple to test the reliability of different groups of experts, using different sources of information (eg either the summary

¹⁹ Note that we are not suggesting that arc uses the "translation" information to inform the awarding of research grants, but to ensure that this information is available for the subsequent granting of translation and partnership awards.

data in Table 3.1 and Table 3.3, or the full case studies in Volume 2) in generating the payback profiles.

The second issue is the definition of the scoring scale. In the current study we used a 1 to 9 scale, where 1 represented no payback and 9 was considered to be the maximum payback that could be expected from such a grant. To aid consistency in scoring it should be possible to generate a series of hierarchical statements that give exemplars of points on the scale. For example, considering the informing policy and product development category, "cited on a local clinical guideline" may be an example of an outcome scoring four points, whereas "cited on a national clinical guideline" (eg Royal College, National Institute for Clinical Excellence (NICE), etc) could score six points. For research leading to industrial product development there would be a different scoring system, for example, the successful development of a novel therapeutic might be an example of an outcome scoring five points, with a therapeutic that had been reviewed and recommended for use by NICE being an example of a seven-point outcome. A variation of this approach has been used in Canada in the development of "language ladders" to assist scientific decision-making in research prioritisation and peer review, and could be a valuable approach in scoring payback profiles.²⁰

The third issue would be to ensure that *arc*'s information management systems are aligned with the collection of payback information. This may require application forms, mid-grant reports, and end-of-grant reports to be revised. Finally, if *arc* did proceed down this route, it would need to be sure that it is cost-effective; ie, that the cost of developing the system and collecting and monitoring the information is no more than the benefit that should be realised from the increased translation of its research.

²⁰ See http://www.ahfmr.ab.ca/grants/docs/2004_03_fellowship_instructions.pdf for an example of the use of language ladders.

1. **Test inter-rater reliability in generating payback profiles.** This will include an assessment of the degree of intra-group and inter-group agreement by different homogenous and heterogeneous groups of experts, and the impact that the presentation of case study information has on scoring.
2. **Development of “language ladders” or “definition trees” for scoring scale.** To aid consistency and reliability in scoring, **arc** could develop a series of hierarchical statements that define the 1 to 9 scale.
3. **Revision of management information systems.** **arc**’s current management information system would need to be reviewed and refined, in order to assure that information on potential payback is captured in grant applications, referees’ reports, committee deliberations, and mid- and end of grant reports.
4. **Assessment of the costs of prospective evaluation versus the benefit of potential increases in translation.** Before implementing a system of prospective monitoring of payback, **arc** needs to be assured of its value for money; that is, that the costs of the system are no more than the potential (but yet to be realised) payback. This would require an audit of the cost of the current peer review and monitoring systems.

Box 4.1: Implementation issues

4.4 Concluding comments

For a number of decades, research-funding organisations have been interested in assessing the payback from their research. In response to this, demand analysts have developed a number of methods to evaluate research. In this study we have combined a number of different approaches to assess the long-term payback from **arc**-funded research in the 1990s. The purpose of the evaluation was to address four specific objectives (repeated in Box 4.2 from Chapter 1, for ease of reading).

Review and document the long-term outcomes of **arc** research grants awarded in the early 1990s.

Identify the factors associated with the translation of research, and begin to develop “early success indicators” that can facilitate the translation of research into practice and can be used to help **arc**’s Development Committee fulfil its remit.

Illustrate the strengths and weaknesses of different modes of research funding, which could inform current practice.

Identify “good news stories” and vignettes of the research process that **arc** could use in its public engagement and fundraising activities.

Box 4.2: Evaluation objectives

As discussed above, we have reviewed and documented the payback from a sample of **arc** research grants that were awarded in the early 1990s. In analysing these paybacks we concluded that committed individuals who are interested in translation are the best “early success indicator” of translation, and we have recommend a number of ways in which **arc**

could stimulate and support that progress. We have explored the impact that different modes of research funded have on translation and have noted that although project grants seemed to provide value for money, **arc** should retain a mixed portfolio of funding. Finally, a number of our case studies provided interesting and informative examples of how research translates into practice and could be used by **arc** in its public engagement and fundraising activities.

We would like to conclude by noting that the rationale for supporting this study was well founded. Research funding agencies such as **arc** need a firm evidence base to support policy and decision-making. The approach, findings and recommendations of this study, we hope, will inform the way in which **arc** “will establish mechanisms ... to stimulate and manage the exploitation of research and educational advances so that they translate into outcomes of practical benefit to people with arthritis”.²¹

²¹ As quoted from the **arc** five-year strategic plan.

REFERENCES

Reference list

Access Economics. 2003. Exceptional returns: the value of investing in health R&D in Australia. Canberra: Australian Society for Medical Research.

Albertini S, Butler J. 1995. R&D networking in a pharmaceutical company: some implications for human resource management. *R&D Management* 25:377-93.

Arthritis Care. 2002. OA nation. Accessed 7 April 2004: <http://oanation.arthritiscare.org.uk/>.

Arthritis Research Campaign (ARC). 2002a. Arthritis: the big picture. Accessed 7 April 2004: http://www.arc.org.uk/about_arth/bigpic.htm.

Arthritis Research Campaign (ARC). 2002b. Research into practice. A five-year strategic plan. Accessed 7 April 2004: http://www.arc.org.uk/about_us/resprac.htm.

Arthritis Research Campaign (ARC). 2004. Research. Accessed 7 April 2004: <http://www.arc.org.uk/research/default.htm>.

Auditor General of Canada. 1994. Report of the Auditor General. Chapter 10: science and technology; management of departmental science and technology activity. Ottawa: Office of the Auditor General of Canada.

Bozeman B, Kingsley G. 1997. R&D value-mapping: a new approach to case study-based evaluation. *Journal of Technology Transfer* 22:33-42.

Buxton M, Elliot R, Hanney S, Henkel M, Keen J, Sculpher M, Youll P. 1994. Assessing payback from Department of Health research and development: preliminary report. Volume 2: eight case studies. HERG Research Report No 19. Uxbridge: HERG, Brunel University.

Buxton M, Hanney S. 1994. Assessing payback from Department of Health research and development: preliminary report. Volume 1: the main report. HERG Research Report No. 19. Uxbridge: HERG, Brunel University.

Buxton M, Hanney S. 1996. How can payback from health services research be assessed? *Journal of Health Service Research and Policy* 1:35-43.

Buxton M, Hanney S. 1997. Assessing payback from Department of Health research and development: second report. Volume 1: the main report. HERG Research Report No. 24. Uxbridge: HERG, Brunel University.

Buxton M, Schneider W. 1999. Assessing the "payback" from AHFMR-funded research. Alberta, Canada: Alberta Foundation for Medical Research.

Buxton M, Croxson B, Hanney S. 1999. Assessing the payback from health R&D: from ad hoc studies to regular monitoring. HERG Research Report No 27. Uxbridge: HERG, Brunel University.

Buxton M, Hanney S, Packwood T, Roberts S, Youll P. 1999. Assessing the benefits from north Thames research and development. HERG Research Report No. 25. Uxbridge: HERG, Brunel University.

Buxton M, Hanney S, Packwood T, Roberts S, Youll P. 2000. Assessing Benefits from Department of Health and National Health Service Research and Development, Public Money and Management, vol 20, no 4, pp 29-34

Buxton M, Hanney S, Jones T. In press. Estimating the economic value to societies of the impact of health research: a critical review. Bulletin of the World Health Organization.

Committee on Science, Engineering and Public Policy (COSEPUP) 1999. Evaluating federal research programs research and Government Performance Results Act. Washington, DC: National Academy Press.

Council on Health Research for Development (COHRED) 2000. Working group on research to action and policy. Lessons in research to action and policy – case studies from seven countries. Geneva: Council on Health Research for Development.

Cronin E, Normand C. 1998. Methodologies for evaluating R&D programmes – a literature review. London: Department of Public Health & Policy, London School of Hygiene and Tropical Medicine.

Croxson B, Hanney S, Buxton M. 2001. Routine monitoring of performance: what makes health research and development different? *Journal of Health Services Research and Policy* 6:226–232.

Dawson G, Lucocq B, Cottrell R, Lewison G. 1998. Mapping the landscape: national biomedical research outputs 1988–1995. London: Wellcome Trust.

Denis J-L, Lomas J. 2003. Convergent evolution: the academic and policy roots of collaborative research. *Journal of Health Services Research and Policy* 8 (supplement 2):1–7..

Drummond M, Davies L, Ferris F. 1992. Assessing the costs and benefits of medical research: the diabetic retinopathy study. *Social Science and Medicine* 9: 973–81.

Ferlie E, Wood M. 2003. Novel mode of knowledge production? Producers and consumers in health services research. *Journal of health Services Research and Policy* 8 (supplement 2):51–7.

Fitch F, Berstein SJ, Dolores Aguilar M, Burnand B, Ramon LaCalle J, Lazaro P, van het Loo M, McDonnel J, Vader JP, Kahan JP. 2001. The RAND/UCLA appropriateness method user's manual, MR-1269-DG-XII/RE. Santa Monica, CA: RAND Corporation.

Fox-Rushby JA. 2002. Disability adjusted life years (DALYS) for decision-making? London: Office of Health Economics.

Gadelha C. 2000. Vaccine research, development and production in Brazil. In: *Lessons in research to action and policy*. COHRED Working Group on Research to Action and Policy. Geneva: COHRED; 9-18.

Garfield E. 1996. How can impact factors be improved? *British Medical Journal* 313:411–13.

Grant J. 1999. Evaluating the outcomes of biomedical research on healthcare. *Research Evaluation* 8:33–8.

Grant J, Allen E. 1999. Evaluating high risk research: an assessment of the Wellcome Trust's Sir Henry Wellcome Commemorative Awards for Innovative Research. *Research Evaluation* 8:201–4.

Grant J, Cottrell R, Cluzeau F, Fawcett G. 2000. Evaluating “payback” on biomedical research from papers cited in clinical guidelines: applied bibliometric study. *British Medical Journal* 320:1107–11.

Grant J, Hanney S, Buxton M. 2004. Academic medicine: time for reinvention. *Research needs researching*. *British Medical Journal* 328:48.

Grimshaw J, Eccles M. 1998. Clinical practice guidelines. In: Silagy C, Haines A, editors. *Evidence-based practice in primary care*. London: BMJ Books.

Hale D, Towse A. 1995. Value of the pharmaceutical industry to the UK economy. London: Office of Health Economics.

Hanney S, Buxton M, editors. 1997. Assessing payback from Department of Health research and development: second report. Volume 1: ten further case studies. HERG Research Report No. 24. Uxbridge: HERG, Brunel University.

Hanney S, Frame I, Grant J, Green P, Buxton M. 2003. From bench to bedside: tracing the payback forwards from basic or early clinical research – a preliminary exercise and proposals for a future study. HERG Research Report No. 31. Uxbridge: HERG, Brunel University. (Available from <http://www.brunel.ac.uk/depts/herg/publications/internal.htm>.)

Hanney S, Gonzalez-Block M, Buxton M, Kogan M. 2003. The utilisation of health research in policymaking: concepts, examples and methods of assessment. *Health Research Policy and Systems* 1 <http://www.health-policy-systems.com/content/1/1/2>.

Hanney S, Soper B, Buxton M. 2003. Evaluation of the NHS R&D methods programme. HERG Research Report No. 29. Uxbridge: HERG, Brunel University.

Hanney S, Mugford M, Grant J, Buxton M. (in press). Assessing the benefits of health research: lessons from research into the use of corticosteroids for the prevention of neonatal respiratory distress syndrome. *Social Science and Medicine*.

Horrobin DF. 1990. The philosophical basis of peer review and the suppression of innovation. *Journal of the American Medical Association* 263:1438–41.

Horrobin DF. 1996. Peer review of grant applications: a harbinger for mediocrity in clinical research? *The Lancet* 348:1293–5.

Innvær S, Vist G, Trommald M, Oxman A. 2002. Health policymakers' perceptions of their use of evidence: a systematic review. *Journal of Health Services Research and Policy* 7:239–44.

Jones T, Hanney S, Buxton M, Burns T. 2004. What British psychiatrists read: questionnaire survey of journal usage among clinicians. *British Journal of Psychiatry* 185:251–7.

Kobelt G, Eberhardt K, Geborek P. 2004. TNF inhibitors in the treatment of rheumatoid arthritis in clinical practice: costs and outcomes in a follow up study of patients with RA treated with etanercept or infliximab in southern Sweden. *Annals of the Rheumatic Diseases* 63:4–10.

Koopmanschap MA, Rutten FFH, van Ineveld BM, van Roijen L. 1995. The friction cost method for measuring indirect cost of disease. *Journal of Health Economics* 14:171–89.

Lavis J, Ross S, McLeod C, Gildiner A. 2003. Measuring the impact of health research. *Journal of Health Services Research and Policy* 8:165–70.

Leong AL, Euller-Ziegler L. 2004. Patient advocacy and arthritis: moving forward. *Bulletin of the World Health Organization* 82:115–20.

Lewison G, Dawson G. 1998. The effect of funding on the outputs of biomedical research. *Scientometrics* 41:17–27.

Lewison G, Devey ME. 1999. Bibliometric methods for the evaluation of arthritis research. *Rheumatology* 38:13–20.

Lewison G, Paraje G. 2004. The classification of biomedical journals by research level. *Scientometrics* 60:145–57.

Lomas J. 2000. Using linkage and exchange to move research into policy at a Canadian Foundation. *Health Affairs* 19:236–40.

McFadzean E. 1998. The creativity toolbox. A practical guide for facilitating creative problem solving. Milton Keynes: Teamtalk Consulting.

Martin B. 1996. The use of multiple indicators in the assessment of basic research. *Scientometrics* 36:343–62.

Martin B, Irvine J. 1983. Assessing basic research: some partial indicators of scientific progress in radio astronomy. *Research Policy* 12:61–90.

Molas-Gallart J, Tang P, Morrow S. 2000. Assessing the non-academic impact of grant-funded socio-economic research: results from a pilot study. *Research Evaluation* 9:171–82.

Murphy K, Topel R, editors. 2003. Measuring the gains from medical research. An economic approach. Chicago, IL: University of Chicago Press.

Mushkin S. 1979. Biomedical research: costs and benefits. Cambridge, MA: Ballinger.

Narin F. 1989. The impact of different modes of research funding. In: Ciba Foundation Conference. The evaluation of scientific research. Chichester: John Wiley & Sons. p 120–40.

Narin F, Pinski G, Gee HH. 1976. Structure of the biomedical literature. *Journal of the American Society for Information Science* 27:25–45.

National Audit Office (NAO). 2003a. Getting the evidence: using research in policymaking (HC 586-1, session 2002–2003). London: National Audit Office.

National Audit Office (NAO). 2003b. An international review on governments' research procurement strategies. A paper in support of getting the evidence: using research in policymaking (HC 586-2, session 2002–2003). London: National Audit Office.

National Institutes of Health. 2000. FY 2000 research program outcomes assessment material. Bethesda, NY: National Institutes of Health.

National Science Board. 1996. Science and engineering indicators – 1996. Washington, DC: US Government Printing Office.

Office of Science and Technology (OST). 2003. PSA target metrics of the UK science base. London: HMSO. (Available at: <http://www.ost.gov.uk>)

Pang T, Sadana R, Hanney S, Bhutta ZA, Hyder AA, Simon J. 2003. Knowledge for better health – a conceptual framework and foundation for health research systems. *Bulletin of the World Health Organization* 81:815-20.

Rosenberg L. 2002. Exceptional economic returns on investments in medical research. *Medical Journal of Australia* 177: 368–71.

Royal Netherlands Academy of Arts and Sciences. 2002. The societal impact of applied research towards a quality assessment system. Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Saklatvala J. 1986. Tumour necrosis factor alpha stimulates resorption and inhibits synthesis of proteoglycan in cartilage. *Nature* 322: 547–9.

Seglen PO. 1997. Why the impact factor of journals should not be used for evaluating research. *British Medical Journal* 314:497.

Smith R. 2001. Measuring the social impact of research. *British Medical Journal* 323:528.

Verhorn C, Landerfeld J, Wagner D. 1982. Measuring the contribution of biomedical research to the production of health. *Research Policy* 11:3–13.

Weisbrod B. 1983. Economics and medical research. Washington, DC: American Enterprise Institute for Public Policy Research.

Wellcome Trust. 2001. Putting NHS research on the map. An analysis of scientific publications in England 1990–97. London: Wellcome Trust.

Wessely S. 1998. Peer review of grant applications: what do we know? *Lancet* 352:301–5.

WK Kellogg Foundation. 2001. Using logic models to bring together planning, evaluation and action. Logic model development guide. Michigan: WK Kellogg Foundation. (Available at <http://www.wkkg.org>)

Wooding S, Grant J. 2003. Assessing research: the researchers' view. RAND document MR-1698-HEFCE. Cambridge: RAND Europe.

Wooding S, Anton S, Grant J, Hanney S, Hoorens S, Lierens A, Shergold M (2004). The returns from arthritis research. Volume 2: case studies. RAND document TR-176-ARC. Cambridge: RAND Europe.

Yin R. 2003. Case study research: design and methods. Third edition. London: Sage.

Yin R, Moore G. 1988. Lessons on the utilization of research from nine case experiences in the natural hazards field. *Knowledge in Society* 1:25–44.

APPENDICES

Appendix A: ranking researchers for case study selection

Summary

We wanted to select our case studies in order to represent the variety of grants awarded by **arc**. We did this by using a case study selection matrix, shown in Table A.1. The design of the matrix is discussed in the main body of the report (in the case study selection section of Chapter 2). For case study selection we generated shortlists for each of the cells of the matrix. The Development Committee then scored the shortlisted grants and made the final selection. This appendix describes how we generated the shortlists, Table A.2 outlines the main stages in generating the shortlists and the issues raised by each stage of the process.

Classifying by funding type

We wanted to classify grants by funding type because the balance between them is a major policy instrument available to **arc**. It was very easy to determine funding type for

Table A.1: Selection matrix

		Basic science	Clinical science	Allied health professional science
Project grants		High success	High success	High success
		Mid-success	Mid-success	Mid-success
Programme grants		High success	High success	High success
		Mid-success	Mid-success	Mid-success
Fellowships		High success	High success	High success
		Mid-success	Mid-success	Mid-success
Institute funding	Kennedy Institute	High success	High success	High success
		Mid-success	Mid-success	Mid-success
	arc Epidemiology Research Unit (arc ERU)	High success	High success	High success
		Mid-success	Mid-success	Mid-success

each grant as the information was held in *arc*'s computerised grants database. With the guidance of the Development Committee we chose to concentrate on the streams of funding that make up the largest fraction of current *arc* expenditure and are those that are most likely to be the subject of future policy decisions: programme funding, project grants, fellowships and institutes.²² Prior to 1992 the two *arc* institutes could apply also for project, programme and fellowship grants in competition with other researchers; for the purposes of this study, this funding was treated also as institute funding.

Classifying by area of work

We expected different areas of work to produce a different spectrum of outputs, so to ensure a reasonable representation of each major area of work we needed to classify grants by area of work. After discussions with key informants and the Development Committee we chose to classify the grants into three categories: basic science, clinical science and allied health professional (AHP) research.²³

arc's grants database contained only grant titles, not grant abstracts. Therefore, instead of manually looking up the abstract for each of the 556 grants in the selection window, we classified grants by the qualifications held by the grant's principal investigator (PI). We used these qualifications, listed in the grants database, to classify PIs as basic scientists, clinical scientists or AHPs. (The list of qualifications used is given at the end of this appendix and a flowchart showing how the PIs were filtered is shown in Figure A.1.) The classification system was not perfect as it did not take account of PIs who, although clinically qualified, had subsequently moved into basic research. We addressed this shortcoming by asking the scientific secretary of *arc* to review our classification. We then reassigned a small number of PIs, based on her knowledge of their research.

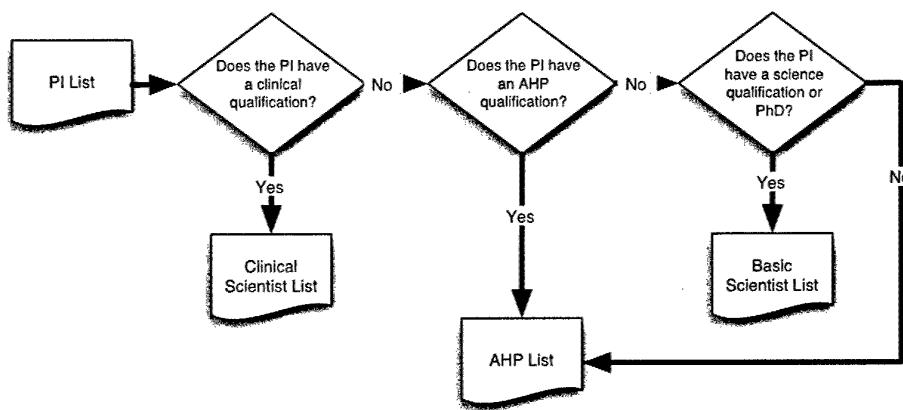


Figure A.1: Method used to classify *arc* researchers into basic scientists, clinicians or allied health professionals

²² Characteristics of the different types of funding are given in Box 2.2.

²³ Work in such areas as physiotherapy and nursing, for which clinical qualifications are not always required.

Table A.2: Outline of method used to generate grant shortlists and issues raised by each stage

Classifying by funding type		Classifying by area of work		Classifying by success	
Reason	Reason				
Major policy instrument	Different areas have different types of outputs Different levels of citation in different areas, so splitting case studies eases classifying by success	To ensure variety of grants examined Research from very high profile researchers may have different routes of translation			
Method	Method	Method	Steps	Scoring papers	Ranking PIs
Use information from arc grant database	Classify grants into clinical, basic or AHP Classify by qualifications of PI from arc grant database Classification reviewed by arc scientific secretary	Use citation measures for PIs to identify "high" impact group and "mid" impact group Identify all PI's papers and add up citation scores Use "High" and "Mid" impact groups to allow comparison between groups	Find core set of papers by filtering by subject and funding acknowledgement Add to core set by looking at co-authorship	Assign "score" to each paper based on Journal Impact Factor (JIF) No perfect method of calculating score, use three alternative techniques	Rank PIs using three different methods of scoring and look for PIs who are suitable in each scoring system
Issues	Issues	Issues	Issues	Issues	Issues
Institutes also received grant funding, these grants were counted as institute funding Focused on most important current types of funding, e.g excluded PhD studentships	PI qualifications may not accurately reflect major differences in areas of research by PI	Large number of grants (550) needs automated system, not possible to look at output of individual grants System does not need to be perfect: needs to ensure variety and allow grouping of case studies by success Using "Low" impact group would give larger difference	Namesakes of the PIs, so system may not identify all papers by PI and may include papers not by PI Large number of PIs requires automated system	Huge number of papers (40,000) means cannot use actual citation numbers, have to use JIF	Will tend to look at "typical" researchers from each group, as excluding those who have large differences between scoring techniques

Classifying by success

We were interested in whether the grants of successful high-profile researchers produced a different spectrum of outputs to those of their less successful colleagues, as this might suggest that different approaches are necessary to ensure the translation of different researchers' work. In order to investigate this, first, we had to be able to compare groups of researchers with differing levels of success. We chose to look at a group of the most successful researchers and a group drawn from the middle of the success ranking. Second, the reason for classifying by success was to ensure that a range of grants were represented in the sample.

Determining the success level of each researcher was the most complex aspect of case study selection. We used each PI's publication record from 1990 to 2002 to assess success: we identified each PI's papers in the ISI Science Citation Index (SCI), and calculated a score for each paper based on the citation level of the journal in which it was published: the Journal Impact Factor (JIF). The following sections describe the background to the technique and each of the three steps.

Background

Bibliometrics is the analysis of publication records and citations. We chose to use it to assess success as it is a widely-accepted, although sometimes controversial, method in quantitative research evaluation. Bibliometrics makes the assumption that "impact" – the number of publications and the number of citations that those publications receive – is a measure of success. We investigated whether the results of bibliometrics (which is relatively easy to apply) correlated with the results from our other, more qualitative investigative tools.

As mentioned above, we wanted to identify two groups of researchers: a "High" and "Mid" impact group. This meant that our classification technique did not need to be perfect; we simply needed to be confident that we had not misclassified "High" researchers as "Mid" and vice versa. Generally, we defined "High" as the top decile of the researchers and "Mid" as the middle tenth (ie from the 45th to 54th percentile); this allowed some degree of uncertainty in rankings before the groups would contaminate one another. We picked the "High" impact group as we hoped that this would be the most likely to have outputs that had been translated. Because of the distribution of bibliometric impact, the majority of case studies are closer to "Mid" impact than "High" or "Low", so we chose "Mid" as our comparison group rather than "Low". Also, we wanted to avoid the potential difficulties of gaining the trust of researchers who felt they had been ranked as "Low".

Success could have been assessed at the level of the researcher or the individual grant. We chose to assess at the level of the researcher for a number of pragmatic reasons:

no computerised records detailing the publication output attributed to each grant exist and it was impractical to assess 556 grants from non-computerised resources such as annual reports;

from previous work we knew that some of the publication outputs from the period of the grant would not be listed in end of year reports;

the end of year reports would not record any publication outputs submitted after the end of the grant was mentioned.

Identifying the papers

We attempted to identify every paper published by each PI between 1990 and 2002. There are two key issues in doing this: comparing across fields and dealing with authors with the same name as our PIs, so-called homonyms or namesakes.

To locate the PIs' papers we searched the SCI, one of the largest databases of bibliometric information for the sciences. We also used the Research Outputs Database (ROD), to provide information on funding acknowledgements on papers with UK author addresses. We picked 1990 as our starting point because it was the beginning of the selection window for our case study grants. As the project was carried out over a number of months we needed to pick a cut-off point for analysis, to ensure that additional citations were not added to the databases between the analysis of different PIs. We used the end of 2002 as a convenient cut-off point. In addition, we used the bibliometric convention of including only papers classified as articles, notes or reviews from peer-reviewed journals.

Database coverage and researcher type

Comparison of bibliometric data across areas of research has two problems: the differing comprehensiveness of the databases and the differences in publication and citation behaviour. The SCI is very comprehensive in basic biomedical science and clinical research. However, its coverage of research in the allied health professions is much less complete, mainly because it does not include a number of journals which are used extensively in these fields.

Publication culture in the allied health professions area also differs greatly from basic and clinical science, with peer-reviewed serial publication seen as far less important. In the AHP field there is more emphasis on "grey" literature (such as reports) and much of the dissemination is via study days and workshops. Because of the limited database coverage and differing publication culture, care must be taken when interpreting the results of the bibliometric analysis for this area of research or comparing it to other areas.

Although there is a similar publication culture in the basic and clinical sciences, the citation culture is often thought to differ, with less citation in the clinical field.

To address these issues when it came to ranking investigators, and determining "High" and "Mid" impact groups, we ranked clinicians, basic scientists and AHPs separately. We classified researchers into these categories as described earlier.

Dealing with homonyms

A recurring problem with bibliometric analysis is determining whether papers are by the researcher of interest or someone else with the same name. We used a novel technique based on identifying an incomplete core set of papers by each PI and then using co-authorship to identify the remaining papers by the PI (outlined in Figure A.2). This section explains the problem in more detail, describes our method and briefly discusses a method that we discarded.

Bibliographic databases, such as the SCI, include a list of authors for each paper. Unfortunately, only the authors' surnames and initials are given. Consequently, some authors may be indistinguishable. For example: Albert Einstein, Angus Einstein and Antonia Einstein would all appear as "Einstein-A". Such namesakes are called homonyms. We needed a way to distinguish the papers by our PIs from their namesakes' papers. In addition, journals may not always use authors' initials in a consistent fashion. For example, Paul Adrien Maurice Dirac may appear in the databases as "Dirac-PAM",

“Dirac-PA” or “Dirac-P”, and on occasion, even as “Dirac-AM”. This further complicates the issue of devising search strategies to find all of an authors papers.

Fortunately, there were no homonyms within our list of arc-funded PIs.

Often, someone with experience of the field can tell whether an author is the PI or a namesake by using other information about the publication, such as the title of the paper, the name of the journal or the author’s address. However, as we had identified over 59,000 publications by searching the databases for our PIs’ names, it was not practical to use this method to check all the papers. To provide us with a way of checking our automated methods, the scientific secretary of arc manually checked the records for 4,706 papers found in a database search for 29 PIs; these PIs were an initial shortlist for case studies generated early in the study. Of the 4,706 papers, only 1,920 papers were scored as being attributable to one of the PIs.

Identifying a core set of papers

Having searched the SCI for every paper by each PI (and their namesakes), we filtered out papers that were not in the field of arthritis and papers that did not acknowledge arc funding – this provided us with a list of papers that we could confidently attribute to the PI.

Subject-based bibliometric filters work by analysing the words of the paper’s title and journal title. To identify arthritis papers we used a filter developed to identify arthritis and rheumatism research, termed the “ARTHR” filter (Lewison and Devey 1999). We determined whether papers acknowledged arc funding using the ROD. This database collated information on funding acknowledgements for papers in the SCI with a UK address.

Comparing this list with the hand-scored list showed (see Figure A.4) that we had identified very few papers that were not attributable to the PIs (a low false positive rate of three papers in 4706), but that we had missed numerous papers that were attributable to the PIs. More importantly, from a ranking point of view, the fraction of papers that were missed varied between PIs (a higher and variable false negative rate, 348 PI papers missed).

Identifying the remaining papers: co-author inclusion

We developed the Co-Author Inclusion (CAI) technique to improve upon the subject and funding filters by identifying the remaining papers by our PIs. CAI identified these papers based on the PIs’ co-authors. The technique can be applied recursively (which we discuss later). The method is described most easily as a number of steps (a flowchart depicting this is shown in Figure A.2).

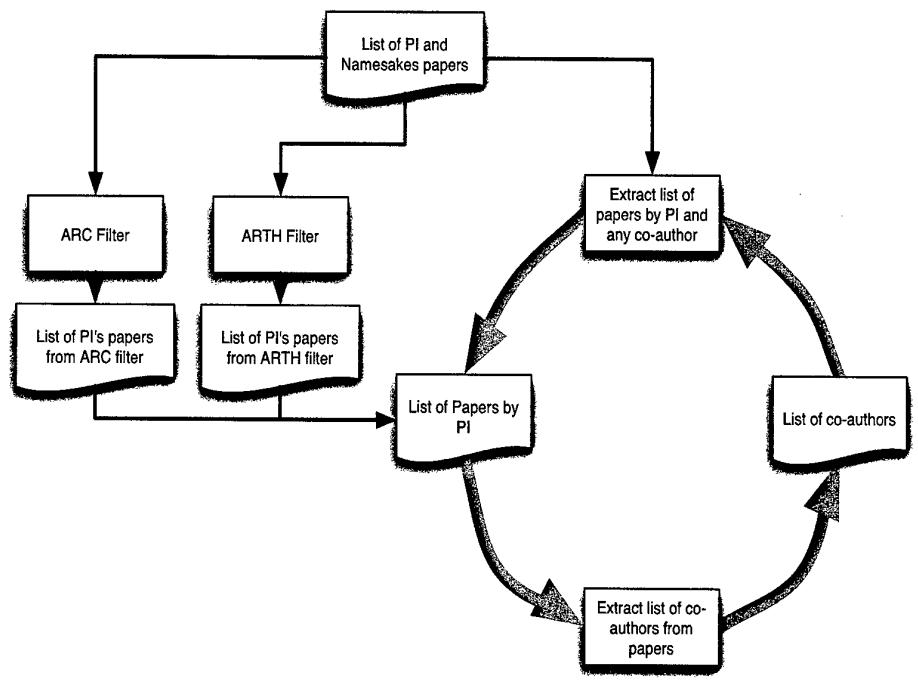


Figure A.2: Flowchart of CAI method

1. Initiate a search on the PI's name and initials in the SCI to generate a complete list of all the papers by the PI and any namesakes.
2. Identify a "seed" set of papers that are known to be authored by the PI. We used the "ARTHR" and "ARC" filters, but we could have used papers that were reported in annual reports.
3. Make a list of all the PIs' co-authors in the "seed" set of papers.
4. Re-examine the original complete list of papers by the PI and namesakes, and assume that any papers including an author on the co-author list are by the PI and not a namesake.

As shown in Figure A.3, using one round of CAI significantly increased the fraction of PIs' papers that we were able to identify. For example, the number of PIs where 100% of their papers are identified doubles from four investigators to eight. However, using CAI does include some papers that are by namesakes; the presence of these false positives are shown by bars where more than 100% of an investigator's papers have been selected. Using one round of CAI, increasing the number of papers falsely identified from three to 40 papers. This is an increase from one investigator to seven investigators with papers falsely credited to them.

Multiple rounds of CAI

The previous section described one round of CAI. As mentioned (and illustrated in Figure A.2) the process can be repeated, taking co-authors from the enlarged list of papers that

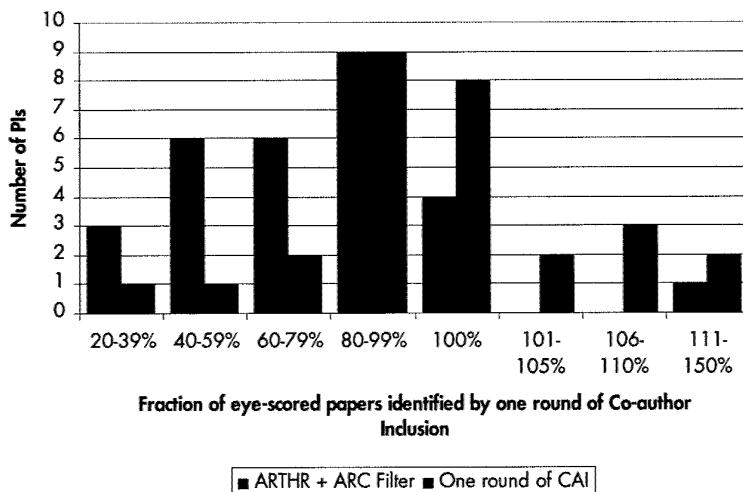


Figure A.3: Percentage of papers found for each PI by two different techniques

resulted from the previous round as the seed list. Note that because every paper must have the PI as an author, the maximum size of the final list of papers produced is limited to all of the papers written by the PIs or their namesakes.

Each round of inclusion may include additional papers by the PI and, in our experience, generally adds fewer papers to the list than the previous round. Each round also has a danger of including false positives in the form of papers from a namesake of the PI, publishing with a namesake of a co-author of the PI.

Because of our relatively small sample of eye-scored papers it is hard to quantify the risk of false positives, but the third round of CAI adds more incorrectly assigned papers than correctly assigned papers (Figure A.4), hence we decided to use two rounds of CAI to generate lists of the publications of each of our PIs.

A worked example of CAI

Figure A.6 shows a worked fictional example that could have come from the world of theoretical physics. Here we are interested in identifying all the papers by Albert Einstein (shown in bold in the figure), and distinguishing these papers from those of Antonia Einstein (shown in plain text). The seed set of papers that we know to be by Albert is highlighted in yellow. From this initial list we are able to identify two of Albert's co-authors: Planck and Heisenberg. We use these authors to add to our list any Einstein-A papers co-authored by Planck or Heisenberg, as we know from our original list that Albert co-authored with these researchers – these papers are highlighted in orange in the second column.

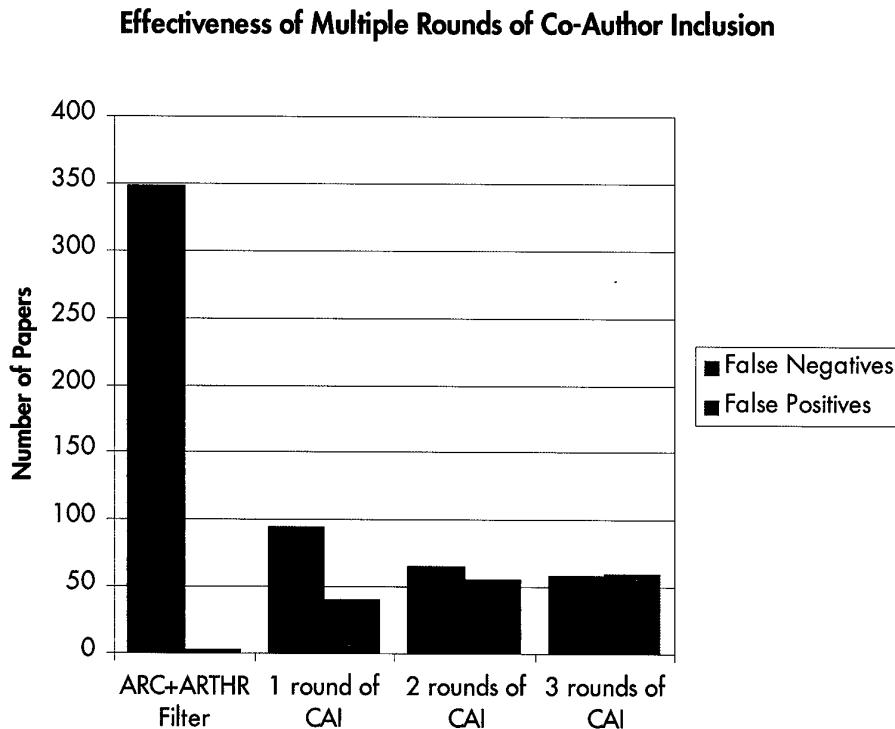


Figure A.4: Comparison of the effectiveness of successive rounds of CAI

We then repeat this process of examining co-authors and discover that Albert also co-authored with Holten and Bohr. We then add any other papers by Einstein-A with either Holten or Bohr to our list. This gives us the list shown in the third column. Finally we repeat the process a third time, adding Schmidt to the list of co-authors and picking up an additional paper by Einstein and Schmidt.

The final column highlights where the technique has worked and where it has failed. This has happened in three cases, all shown in red; and these demonstrate the three failure modes of the technique.

In the first, a single author paper by Albert, there are no co-authors to allow us to determine whether this Einstein is Albert or not.

In the second case, a paper by Einstein and Dirac is the only paper in the set that Albert co-authored with Dirac so we have not picked him up as a co-author – this is likely to be an increasing problem as the size of the seed set is decreased.

In the third case, we have a paper by Einstein and Schmidt that was included even though it was by Antonia. This has happened because a Schmidt O is a common name for a co-author and both Antonia and Albert have authored with a Schmidt O.

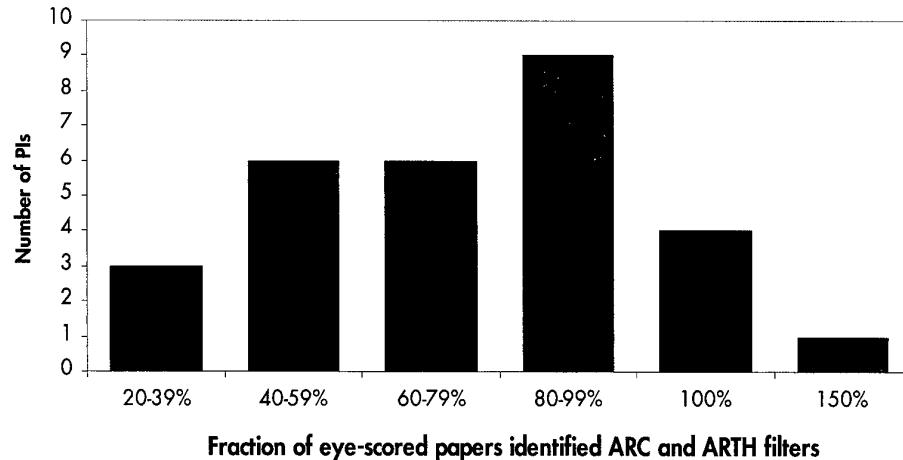
Full List of Papers	First Round of CAI	Second Round of CAI	Third Round of CAI
1 Einstein A, Planck M	Einstein A, Planck M	Einstein A, Planck M	Einstein A, Planck M
2 Einstein A	Einstein A	Einstein A	Einstein A
3 Einstein A, Planck M, Bohr N	Einstein A, Planck M, Bohr N	Einstein A, Planck M, Bohr N	Einstein A, Planck M, Bohr N
4 Einstein A, Planck M, Holten F	Einstein A, Planck M, Holten F	Einstein A, Planck M, Holten F	Einstein A, Planck M, Holten F
5 Einstein A, Bohr N	Einstein A, Bohr N	Einstein A, Bohr N	Einstein A, Bohr N
6 Einstein A	Einstein A	Einstein A	Einstein A
7 Einstein A, Hastings M	Einstein A, Hastings M	Einstein A, Hastings M	Einstein A, Hastings M
8 Einstein A, Heisenberg WK	Einstein A, Heisenberg WK	Einstein A, Heisenberg WK	Einstein A, Heisenberg WK
9 Einstein A	Einstein A	Einstein A	Einstein A
10 Einstein A, Dacre P	Einstein A, Dacre P	Einstein A, Dacre P	Einstein A, Dacre P
11 Einstein A, Heisenberg WK	Einstein A, Heisenberg WK	Einstein A, Heisenberg WK	Einstein A, Heisenberg WK
12 Einstein A, Heisenberg WK, Planck M	Einstein A, Heisenberg WK, Planck M	Einstein A, Heisenberg WK, Planck M	Einstein A, Heisenberg WK, Planck M
13 Einstein A, Dirac PAM	Einstein A, Dirac PAM	Einstein A, Dirac PAM	Einstein A, Dirac PAM
14 Einstein A, Bohr N, Smith J	Einstein A, Bohr N, Smith J	Einstein A, Bohr N, Smith J	Einstein A, Bohr N, Smith J
15 Einstein A, Smith J	Einstein A, Smith J	Einstein A, Smith J	Einstein A, Smith J
First Co-Author List	Second Co-Author List	Third Co-Author List	
Planck M	Planck M	Planck M	
Heisenberg WK	Heisenberg WK	Heisenberg WK	
	Holten F	Holten F	
	Bohr N	Bohr N	
		Smith J	

Figure A.5: A worked example of CAI

Rejected techniques

Initially, we had considered using just the “ARTH” and “ARC” filters to identify the PIs’ papers; however, the problem with these techniques was that while they were good at excluding papers by the PIs’ namesakes from our lists, they tended to miss papers that were by the PI (see Figure A.6). This occurred for two reasons. First, much of the research that is funded by arc investigates the basic biochemical and biological processes underlying arthritis and rheumatism. These processes are common to other disease states and indeed normal tissue, therefore the papers reporting the research often have no obvious link to arthritis. Second, in the early 1990s there was much less fastidiousness in acknowledging funding, so many papers that were supported by arc may not acknowledge this support.

Another possible technique to identify the papers of our PIs would have been to use addresses. Bibliometric databases contain information on the address of all the authors of each paper. We could have used the address of the PIs when they applied for their grants, from arc’s database, to allow us to filter for only papers from that address. Provided that namesakes of the PI were not based at the same address, we could filter them out successfully. For example, we could look only for papers published by A Einstein that originated in Zurich. Any such filtering method would have to allow for the movement of researchers during the 12-year window of analysis. This was always likely to be a complex technique and we would need a method to deal with the considerable number of PIs who changed their addresses between 1990 and 2001. However, it proved impossible to apply this technique as the information on author addresses available in the SCI was not consistent enough, for example: St Bartholomew’s Hospital, London might appear as St Bartholomews, St Barts, Barts Hospital, or St Barts Hosp. Given the number of researchers concentrated in London and other centres, a location-based system needed to be able to



PIs with more than 100% of papers have papers incorrectly assigned to them. Comparison is to eye-scored set. Only one PI has more papers than they should have (150% column), this excess represents three papers.

Figure A.6: Number of PIs and the fraction of their papers missed by ARC and ARTH filters

distinguish between institutional addresses. We could not use postcodes as an alternative way of determining address, as information on postcodes from the SCI is not reliable for the early 1990s. Furthermore, if researchers has gone abroad for sabbaticals we would have excluded the papers that they published from abroad.

Scoring papers

Having identified each PI's papers we needed to score each paper in order to rank the PIs by total publication score. There is no perfect method of determining scores for papers and this section discusses the three methods that were used. There are two stages to this process: determining how much each paper should score and determining how much of that score should go to each author.

Various methods can be used to assign scores to papers; unlike finding the complete list of papers for a PI, there is no single correct method. We have adopted the "triangulation" approach: we have used a number of different techniques and the similarity of the results to give us an idea of the confidence that we should have in our rankings.

As mentioned earlier, a problem with scoring papers by number of citations is the difference in citation behaviours between fields: some fields such as genetics cite very highly, while others such as nursing have much lower citation levels. We have tried to avoid the largest discrepancies this issue causes by dividing the PIs into broad types: basic scientists, clinical scientists and AHPs (as explained earlier).

$(C_{0-4})_i$ is average number of citations for the journal in which paper i is published over the five years from the publication date

A_i is number of authors for paper i

$$\text{Total Journal Impact Factor: } \sum_{i=1}^n (C_{0-4})_i$$

$$\text{Total Scaled Journal Impact Factor: } \sum_{i=1}^n \log_{10}(1 + (C_{0-4})_i)$$

$$\text{Total Fractionated Journal Impact Factor: } \sum_{i=1}^n \frac{(C_{0-4})_i}{A_i}$$

Figure A.7: Equations used to calculate PI publication scores

Journal impact factor

The average number of citations received by each paper in a journal, over the five years from publication, is referred to as the five-year JIF or C_{0-4} . The JIF is the number of citations that one might expect a paper published in the journal to receive in the five years following its publication. For our first technique we used this JIF as the score for each paper, and assigned the whole score to each author. We used JIFs calculated for each two-year span of the time window, eg 1990–1991, 1992–1993, etc.

Scaled JIF

JIF has a very wide range and the exact relationship between the number of citations and the importance of a paper is unclear. In general, the more frequently-cited papers are considered to be more important, but a question remains as to whether this is a linear relationship: is a paper that is cited 100 times actually 20 times more valuable than a paper that is cited five times? The wide range of JIF means that if the relationship between JIF and importance is linear, then one paper in *Nature* (JIF₁₉₉₈ = 100.6) is equivalent to approximately 12 papers in *Annals of the Rheumatic Diseases* (JIF₁₉₉₈ = 8.54). An alternative technique is to use a logarithmic relationship between JIF and importance. This has the effect of reducing the significance of publication in journals with a high JIF. This was the second technique that we used, again assigning the whole score to each author.

Fractionated JIF

All the authors can be given full credit for each paper, or the score can be divided between them. In the two examples considered so far, each author has received the full credit for each paper: if the paper scored 2.4, each author gets 2.4. This does not reflect the fact that different authors will have made different contributions to a paper. In the biological sciences it is often the case that the first author will have made the greatest contribution, whereas the last author may have supervised the work. However, personal and political issues also may affect the order of authorship. Without contextual information it is impossible to allocate scores at this level of detail, however, it is often assumed that as the number of authors increases, the credit that they are allocated for a paper should decrease. To reflect this, in our final approach we used JIF as the score for each paper but then

divided this score equally between all the authors – so if a paper had two authors and a JIF of 2.4, each author would get 1.2.

Examining the rankings

The final stage of shortlisting our PIs was to generate rankings to allow us to select “High” and “Mid” impact case study candidates for each cell of the matrix.

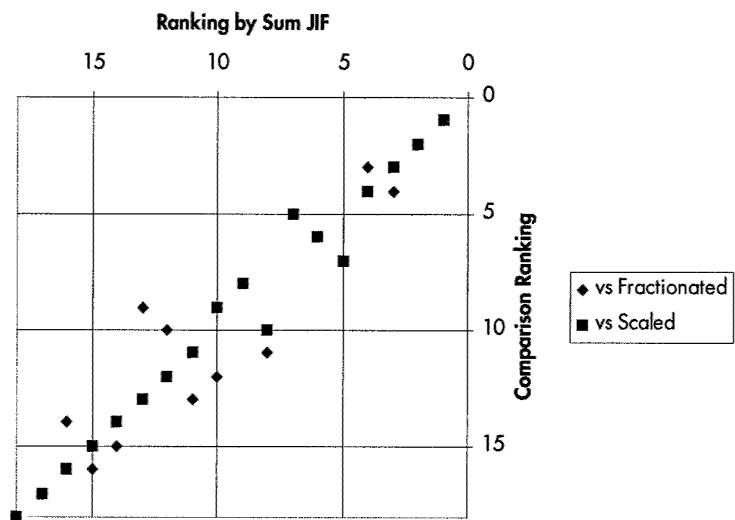
Different methods of scoring provide similar rankings

Changing how papers score and how this score is divided between authors has little effect on the ranking of most researchers, as illustrated in Figure A.8. This means that distinguishing between a cadre of high-impact researchers and mid-impact researchers was relatively easy: there were few researchers who would be “High” impact (generally top decile) by one scoring system and “Mid” impact (middle tenth) by another. This suggests that most of the PIs have similar publication habits.

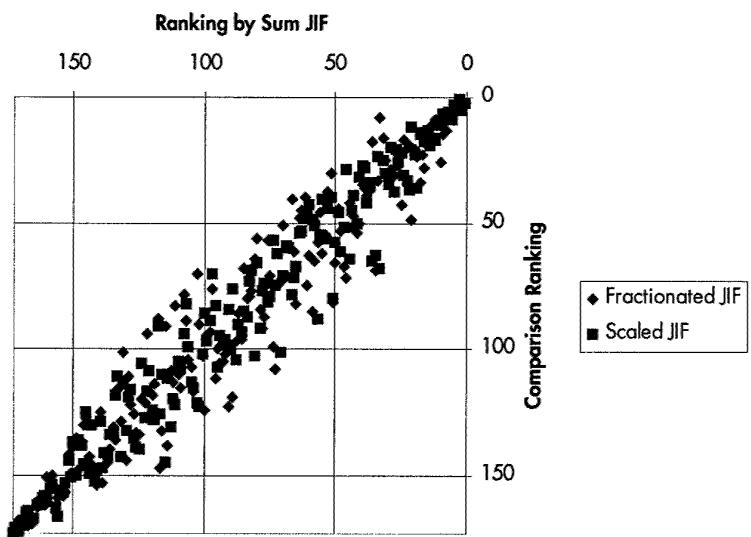
Variation in ranking is highest for mid-impact researchers

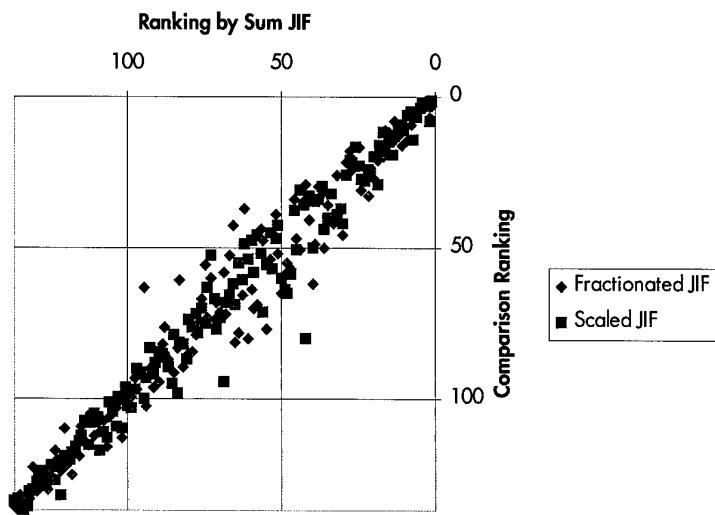
It is noticeable from Figure A.9, which shows the variation in ranking produced by different scoring systems, that the PIs in the centre of the ranking have more variation between ranking systems than the ones at the extremes. This is explained partially by looking at the distribution of scores among the researchers. The distribution of scores for summed JIF of basic scientists is shown in Figure A.9; other researcher types and methods of scoring show the similar distributions. The variation between adjacent PIs in the ranking is far lower in the centre section of the ranking than between the higher-ranked researchers, so small changes in score (for example, by changing the scoring method) have a larger effect on ranking than they do at the top end of the ranking.

Allied Health Professionals Comparison Ranking



Basic Scientist Comparison Ranking



Clinical Researchers Comparison Ranking

(See Figure A.7 for equations used)

Figure A.8: Comparisons between ranking when score is determined by summation of JIF, fractionated JIF or scaled JIF

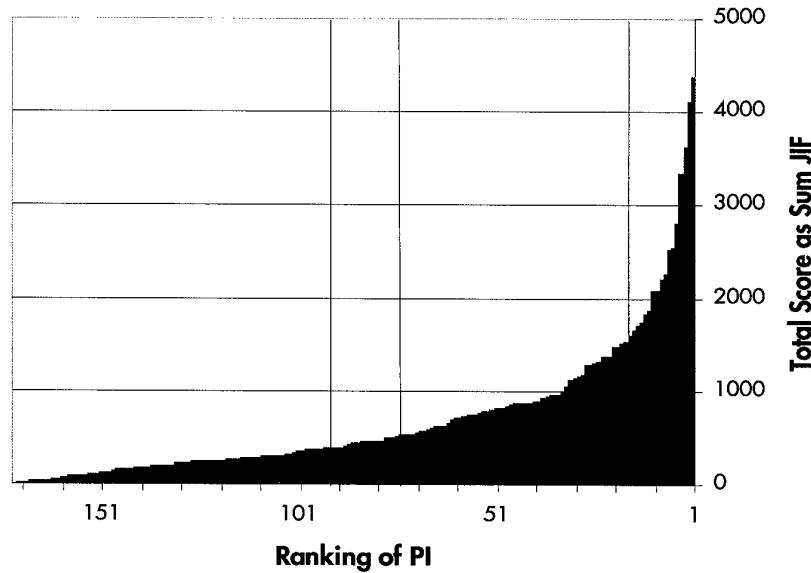
Comparison of Ranking and Score for Basic Science Researchers

Figure A.9: Relationship between ranking of researchers and total bibliometric score (as sum of JIF) for basic scientist PIs

Closing remarks

Categorising grants by funding type and area of research proved relatively straightforward, thanks to the information provided in arc's grants database. Categorising PIs by success provided a much more complicated and involved procedure than we had expected. However, during the course of testing the methods used, we have developed a novel and robust approach to assessing the bibliometric success of a large number of researchers. This would make it much simpler to apply the methodology in a future study.

Qualifications used for classification

Qualifications leading to classification as clinical researcher

BAO (Bachelor of Obstetrics), BChBAO (Bachelor of Surgery and Bachelor of Obstetrics), BChir (Bachelor of Surgery), BMedSci (Bachelor of Medical Sciences), BS (Bachelor of Surgery), ChB (Bachelor of Surgery (Scotland)), DCH (Diploma of Child Health), DM (Doctor of Medicine), DSMSA (Diploma of Sports Medicine, Society of Apothecaries), FACC (Fellow of the American College of Cardiology), FFD (Dental qualification), FMGEMS (Foreign Medical Graduate Examination), FRACP (Fellow of the Royal College of Australasian Practitioners), FRC (From "FRC PATH" counted as FRC, ignored under PATH), FRCGP (Fellow of the Royal College of General Practitioners), FRCOG (Fellow of the Royal College of Obstetrics and Gynaecology), FRCP (Fellow of the Royal College of Physicians), FRCP(E) (Fellow of the Royal College of Physicians (Edinburgh)), FRCP(UK) (Fellow of the Royal College of Physicians (UK)), FRCPPath (Fellow of the Royal College of Pathology), FRCPCH (Fellow of the Royal College of Paediatrics and Child Health), FRCPE (Fellow of the Royal College of Physicians (Edinburgh)), FRCS (Fellow of the Royal College of Surgeons), FRCS(Orth) (Fellow of Royal College of Surgeons (Orthopaedic)), FRCSE (Fellow of the Royal College of Surgeons (Edinburgh)), FRCSEI (Fellow of the Royal College Surgeons (Ireland)), HonMRCP (Honorary Member of the Royal College of Physicians), LRCP (Licenciate of the Royal College of Physicians), MB (Bachelor of Surgery), MBBCh (Bachelor of Surgery), MBBChir (Bachelor of Medicine and Bachelor of Surgery), MBCB (Bachelor of Surgery), MD (Medical Doctor), Med (From "soc Med", counted as "Med" ignored as "soc"), MRCGP (Member of the Royal College of General Practitioners), MRCP (Member of the Royal College of Physicians), MRCP(UK) (Member of the Royal College of Physicians), MRCPPath (Member of the Royal College of Pathologists), MRCPI (Member of the Royal College of Pathologists (Ireland)), MRCS (Member of the Royal College of Surgeons), MS (Master of Surgery), RCS (Royal College of Surgeons), RCSI (Royal College of Surgeons (Ireland)).

Qualifications leading to classification as an allied health professional researcher (if not already classified as clinical researcher)

BSocPsy (Bachelor of Social Sciences), DipClinPsychol (Diploma in Clinical Psychology), DipCOT (Diploma of the College of Occupational Therapists), FFCM (Fellow of the Faculty of Community Medicine), FFPHM (Fellow of the Faculty of Public Health Medicine), MCSP (an AHP qualification), MFPHM (Membership of the Faculty of Public Health Medicine), MMACP (Membership of the Manipulation Association of Chartered Physiotherapists), SROT (State Registered Occupational Therapist).

Qualifications leading to classification as a basic scientist (if not already classified as clinical scientist or AHP researcher)

AMIMechE (Associate Member of the Institute of Mechanical Engineering), BA (Bachelor of Arts), BE (Bachelor of Engineering or Bachelor of Education), BSc (Bachelor of Sciences), BSc(Eng) (Bachelor of Sciences (Engineering)), BVetMed (Bachelor of Veterinary Medicine), BVMS (Bachelor of Veterinary Medicine), CEng (Chartered Engineer), CBiol (Chartered Biologist), CEng (Chartered Engineer), DEng (Doctor of Engineering), FDS (From "FDSRCS" Fellow of Dental Surgery, Royal College of Physicians, count under FDS), FDSRCS (Fellow of Dental Surgery of the Royal College of Physicians), FIBiol (Fellow of the Institute of Biology), FIM (engineering qualification), FIMechE (Fellow of Institute of Mechanical Engineering), GRSC (Graduateship of the Royal Society of Chemistry), MIBiol (Member of the Institute of Biology), MRCVS (Member of the Royal College of Veterinary Surgeons), VMD (Doctor of Veterinary Medicine).

Qualifications not considered diagnostic

MRPharmS (Member of the Royal Pharmaceutical Society), DPhil (Doctor of Philosophy), DSc (Doctor of Science), FRS (Fellow of the Royal Society), MPhil (Master of Philosophy), MSc (Master of Sciences), PhD (Doctor of Philosophy), ScD (Doctor of Science), Ed(Orth) (From FRCS Ed(Orth) counted as FRCS), FMedSci (Fellow of Medical Sciences), FRCCC (unknown), FRSE (Fellow of the Royal Society (Edinburgh)), AFBPS (Associate Fellow of the British Psychological Society), AKC (Associate of King's College), BCH (Bachelor of Surgery), BPharm (Bachelor of Pharmacy), CBE (Commander of the Order of the British Empire), Cert (From "Cert SAO", meaning: Certificate of Small Animal Orthopedics, count under Cert), DCR (Unknown), DDSc (Doctor of Dental Science), DIC (Diploma of Imperial College), FIBMS (Fellow of the Institute of Biomedical Sciences), FREng (Fellow of the Royal Society of Engineering), ILTM (Institute of Learning and Teaching), MA (Master of Arts), MA(Oxon.) (Master of Arts), OBE (Officer, Order of the British Empire).

Appendix B: case study project descriptions

Case study A

A randomised controlled trial (RCT) to evaluate the effectiveness, in terms of clinical outcomes and costs, of an exercise programme to encourage normal activities in a community setting for patients with low back pain.

Case study B

The research project assessed occupational activity as a risk factor for osteoarthritis (OA) of the hip, using a case control epidemiological design.

Case study C

The research project investigated the basic biochemistry of proteins involved in destruction of cartilage and other connective tissue in joints. The grant focused on two proteins: Tissue Inhibitors of Metallo Proteases 1 and 2 (TIMPs 1 and 2). These proteins were known to inhibit Matrix Metallo Proteases (MMPs), proteases that play an important role in cartilage destruction.

Case study D

The project investigated the mechanisms involved in causing, maintaining and ending inflammation in reactive arthritis. When caused by chlamydia, the condition is known as sexually-acquired reactive arthritis (SARA). The study aimed to investigate the role of dendritic cells in SARA; in particular to examine whether dendritic cells were continuing to present bacterial peptides after they had moved from the site of infection to the joints.

Case study E

The aim of the project was to understand how different aspects of knee OA interrelate. The central hypothesis was that different tissues and processes are involved to a varying degree in different patients, explaining the heterogeneity of outcome and implying that different approaches to prevention and therapy are possible. This involved a combination of clinical and basic research.

Case study F

In the late 1980s, the PI's lab had succeeded in cloning the human type X collagen gene (collagen X). For the programme grant, they proceeded to characterise the gene and its promoters, as well as to try and identify the mutations that could cause certain types of dysplasias.

Case study G

The aim of the research was to delineate the basic mechanisms involved in immune complex processing by the mononuclear phagocytic system. This would help to explain which processing factors might be responsible in the pathogenesis of systemic lupus erythematosus (SLE), a rare and poorly understood rheumatic disease.

Case study H

The research grant assessed the hypothesis that the apparent down-regulation of immune responses in rheumatoid arthritis (RA) may be a perfectly normal response to chronic inflammation in a specific environment. The role of T cells (types of white blood cell which are normally very effective at defending the body against disease) in this process is crucial, but unclear.

Case study I

The research tested the hypothesis that muscle strength and the decomposition of OA are related. The objectives of the research were to characterise the pattern of weakness in patients with knee OA and to determine whether reduced muscle strength correlates with presence of OA or primarily with symptomatic OA. In addition, one of the aims of the junior clinical research fellowship was to provide training in research skills.

Case study J

The research examined the role of the anterior cruciate ligament as contributing an important proprioceptive role in the provision of stability of the knee joint. The aim of the research was to measure proprioceptive changes in the surgically-repaired cruciate deficient knee and to evaluate the efficacy of two different physiotherapy programmes. Proprioceptive refers to the ability to sense the position, location and orientation of the knee.

Case study K

Antiphospholipid syndrome (APS) is a disease with a similar pathogenesis as SLE. This disorder can be asymptomatic and may only manifest itself during pregnancy, where it is characterised by recurrent miscarriages. Various therapies had been suggested for prevention of foetal death in APS. The aim of the study was to undertake a RCT of low-dose aspirin plus low-dose subcutaneous heparin in women with APS.

Case study L

The selected strand of research supported by the institute's core grant aimed to investigate the activity of specific cytokines in RA. Specifically, the study tested, through a small clinical trial and subsequent larger multi centre trial, the hypothesis that RA could be treated by using targeted Tumour Necrosis Factor alpha (TNF α) inhibitors.

Case study M

RA is a multifactorial disease to which there is a significant component, about one-third of which can be accounted for by human leukocyte-associated antigen (HLA)-linked genes. The aim of the research was to generate a linkage map of genetic effects in RA in 130 nuclear families containing affected sibling pairs. Ultimately, this type of study should lead to identifying genes that influence the development of RA.

Case study N

The research under consideration aimed to improve the understanding of healthy cartilage and the mechanisms of its degeneration in OA. The work consisted of two strands. The first was to investigate changes in the proteoglycan aggrecan, which is of major importance to cartilage function. The second strand studied the enhanced activity of chondrocytes in OA in culture models, with the aim of throwing light on the molecular and cellular mechanisms that cause cartilage damage.

Case study O

The project aimed to identify and characterise, through linkage analysis, non-HLA RA susceptibility genes, and to evaluate their relative contribution to the aetiology of the

disease. The research under consideration was awarded originally as a postgraduate research fellowship which was then rolled into the institute core funding.

Case study P

The Norfolk Arthritis Register (NOAR) was set up in 1990 as a means to study new case of inflammatory arthritis as they occur in the community, and to follow patients prospectively in order to investigate the natural history of the condition. NOAR began with three aims:

1. to establish the incidence of RA;
2. to look for evidence of clustering of RA; and
3. to look for early predictors of the disease.

There is increasing pressure for research funders to demonstrate, and seek to maximise, the payback from the research they fund. This report, prepared for and funded by the Arthritis Research Campaign (arc), presents the results of an evaluation of 16 research grants awarded by arc in the early 1990s. The main objective was to develop a system for evaluating arthritis research, with a view to allowing arc to stimulate and manage the exploitation of research advances so that they translate into outcomes of practical benefit to people with arthritis.

The report presents a framework that conceptualises the relationship between research inputs, process, output and outcomes. Using this framework, we catalogue a diverse range of research output and outcomes arising from these 16 grants and make a series of quantitative and qualitative assessments comparing, for example, payback from project grants versus programme grants. In conclusion, we make six observations:

- There is a diversity of research payback.
- The researcher is the key driver of research translation.
- Short, focused project grants seem to provide value for money.
- Intended and unintended flexibility in funding is used advantageously.
- Referees' contributions to the peer-review process are of variable benefit.
- The payback framework could be operationalised and embedded by arc.

The companion Volume 2 is a collection of the case studies. These case studies all follow a similar format based on the conceptual model and provide a rich and detailed narrative on the payback of each research grant.

Both volumes of the report are available, as PDF files, from <http://www.rand.org>.

This product is part of the RAND Corporation monograph series. RAND monographs present major research findings that address the challenges facing the public and private sectors. All RAND monographs undergo rigorous peer review to ensure high standards for research quality and objectivity.

